ANA CRISTINA DA CUNHA FIGUEIREDO DE CARDOSO LIMA

Exploring the regulatory networks of spermatogenesis: from candidate gene approaches to integrative functional genomics

Tese de Candidatura ao grau de Doutor em Biologia Básica e Aplicada submetida ao Instituto de Ciências Biomédicas Abel Salazar da Universidade do Porto.

Orientador – Doutora Alexandra Manuel Ferreira Lopes

Categoria – Investigadora

Afiliação – Instituto de Investigação e Inovação em Saúde, Universidade do Porto (I3S); Instituto de Patologia e Imunologia Molecular da Universidade do Porto (Ipatimup), Porto, PORTUGAL.

Coorientador – Doutor Donald Franklin Conrad

Categoria – Assistant Professor

Afiliação – Washington University in St. Louis - School of Medicine, St. Louis, MO, USA

Coorientador - Professor Doutor António Amorim dos Santos

Categoria – Professor Catedrático

Afiliação – Instituto de Investigação e Inovação em Saúde, Universidade do Porto (I3S); Instituto de Patologia e Imunologia Molecular da Universidade do Porto (Ipatimup); Faculdade de Ciências da Universidade do Porto, Porto, PORTUGAL

Research work coordinated by:

Apoio financeiro

Projecto

Este trabalho foi parcialmento financiado pela Fundação Portuguesa para a Ciência e Tecnologia FCT/MCTES (PIDDAC) e co-financiado por fundos Europeus (FEDER) através do programa COMPETE (PTDC/SAU-GMG/101229/2008 e IF/01262/2014, atribuídos a AML), CIGMH e Pest-OE/SAU/UI00009/2011. IPATIMUP é um laboratório associado do Ministério Português de Ciência, Tecnologia e Ensino Superior e parcialmento suportado pela FCT.

Este trabalho foi ainda financiado pelos National Institutes of Health (NIH),USA, através dos projectos R01HD078641, R01MH10810 e R01HG007178, atribuidos a DFC.

Canditato

Atribuição de uma bolsa indivudual de Doutoramento da Fundação Portuguesa para a Ciência e Tecnologia (FCT- (SFRH/BD/51695/2011) e apoio financeiro para deslocações aos Estados Unidos da America proporcionado pela Fundação Luso-Americana para o Desenvolvimento (FLAD).

Ao abrigo do art.º 8º do Decreto-Lei n.º 388/70, fazem parte integrante desta dissertação os seguintes manuscritos já publicados, em revisão ou em preparação:

Lima, A.C; and Lopes, A. M. (June 2014) Autosomal Mutations and Spermatogenic Failure. In: eLS. John Wiley & Sons, Ltd: Chichester. DOI: 10.1002/9780470015902.a0025310

Lima, A.C., Carvalho, F., Goncalves, J., Fernandes, S., Marques, P.I., Sousa, M., Barros, A., Seixas, S., Amorim, A., Conrad, D.F., Lopes, A.M. (2015). Rare double sex and mab-3-related transcription factor 1 regulatory variants in severe spermatogenic failure. *Andrology* **3**: 825-833.

Seabra, C.M., Quental, S., Lima, A.C., Carvalho, F., Goncalves, J., Fernandes, S., Pereira, I., Silva, J., Marques, P.I., Sousa, M., Barros, A., Seixas, S., Amorim, A., Lopes, A.M. (2015) The mutational spectrum of WT1 in male infertility. *The Journal of urology* **193**: 1709-1715.

Lima, A.C.*, Jung, M.*, Rusch, J., Usmani, A., Lopes, A.M., Conrad, D.F. Multispecies purification of testicular germ cells*. Biology of Reproduction* (***in revision***)

Lima, A.C., Yuan, N.H.R., Amorim, A., Conrad, D.F., Lopes, A.M. Challenges and solutions for ribosome profiling with limited cell numbers: the example of murine male germ cells. (***in preparation***)

Em cumprimento do disposto no referido Decreto-Lei, a candidata declara que participou na obtenção, análise e discussão dos resultados, bem como na elaboração das publicações, sob o nome Lima, A.C.

***** *autoria partilhada*

Agradecimentos

Doing a PhD is a huge challenge both in terms of work and on a personal level. I usually call it my personal roller coaster….with so many ups and downs, sometimes in a matter of hours. It is therefore an adventure that brings you amazing work skills and shows you a great deal about yourself and what you are capable of. However, it is not something you do alone…you require a large network of support because you will need it sooner rather than later. Here are my acknowledgments to MY network…my rocks!

À **Alexandra**, minha orientadora: Por onde começar? Todos temos pessoas que mudaram a nossa vida para sempre, e para mim, a Alexandra é sem dúvida uma delas. Tudo que possa dizer não chega para exprimir a minha gratidão, que é eterna! Por todo o apoio, motivação e conhecimentos que me transmitiu ao longo desta jornada científica, por todas as discussões e ideias partilhadas, por ser um modelo a seguir...! Por estar sempre lá, nos momentos bons e sobretudo nos mais difíceis. Por saber que pode ser sábado à noite ou domingo de manhã mas se eu estiver confusa ou perdida sem saber o caminho a seguir, ela está sempre à distância de um telefonema com o seu sorriso e o seu talento de, numa única conversa, conseguir organizar as minhas ideias e repôr ordem ao caos. Pelos momentos fantásticos nas nossas viagens e por me mostrar a importância de viver novas experiências e desafios. Obrigada não chega para agradecer tudo que ganhei com a nossa amizade e colaboração durante estes anos. Serás sempre um pilar no meu progresso científico e pessoal.

To **Don**, my co-advisor: For the kindness that I received when joining the group. All the valuable insights and work on the development of this project. Especially for the constant motivation, believing in me more than myself and for teaching me how to celebrate the small achievements, always showing me the progress we were making in times that all I could see were the problems ahead. For his genuine concern about me as a person, my scientific career and my integration in St. Louis. For being an amazing PI that cultivates group unity both in terms of research and social events, creating a fantastic lab environment to grow as a scientist and as an individual.

Ao **Professor Amorim**, meu coorientador: Por todas as discussões científicas, conhecimento que me transmitiu, pela constante presença e apoio ao longo destes anos e pelos momentos partilhados nos convívios do grupo. Agradeço sobretudo estar sempre disponível para qualquer ocasião, mesmo quando isso implicava tratar de assuntos para “ontem”. O grupo da

Genética Populacional foi uma famíla na qual me senti sempre incluída durante todos estes anos a agradeço-o por sempre ter lutado para que assim o fosse!

Ao GABBA, por me ter dado esta fantástica oportunidade! À Catarina Carona por tornar tudo isto possível e estar sempre lá para nós. À Professora Maria de Sousa, sempre uma inspiração, e ao Professor Alexandre do Carmo. A todos os GABBAs, pelos momentos ciêntíficos e de convívio ao longo destes anos. Ao GABBA 15th por todos os desafios que ultrapassámos juntos e tudo que aprendemos na nossa jornada, em especial à Danica e ao Dário.

I would also like to thank all the co-authors for their valuable collaboration that made this work possible.

A todo o grupo da Genética Populacional do IPATIMUP/I3S, especialmente ao Rui, Sandra, Sofia e Luís por todo o apoio, disponibilidade e companheirismo. Aprendi muito e levo ainda muitas memórias queridas destes anos que partilhei convosco.

Às minhas meninas do Ipas, em especial à Pati e à Cats, VC e Ritinha, companheiras de lab e de desatinos! Agradeço todas as discussões ciêntificas, a ajuda e a amizade, o carinho e a paciência que sempre tiveram comigo. À Tati por permitir que lhe transmitisse os meus conhecimentos, mostrando-se sempre disposta a ajudar e a aprender!

To ALL the members from the Conrad Lab at WASHU! I am grateful for all the help, at work or otherwise, and for creating an amazing work environment! For the great lab outings and for making me feel at “home” away from home…

To Liina, Maya and Amy… I’m going to make it short: THANK YOU! You know how grateful I am for having your friendship. To all my friends in St. Louis for so many special moments!

À Ju e ao Bis, por me aturarem à mais anos do que tenho memória! Obrigada por mesmo longe estarem sempre perto!

To Mathias, for being one of the most amazing people I know! For his companionship, the constant concern with me and my work, for accepting the challenge of adopting two dogs and all the responsibility that comes with it. For keeping my sanity, making me grow and being, above all, my BEST friend!

Aos MEUS! Foram muitos os desafios até chegar aqui e sem a minha **FAMÍLIA** nunca o teria conseguido. Um obrigado de coração à minha Mãe por ser tudo e mais do que alguém possa

desejar, sem dúvida mais do que sou digna de merecer! Ao Jorge pelo carinho e apoio incondicional, ao meu irmão e cunhada por me darem uma princesinha que alegra os meus dias! Ao meu Pai por toda ajuda e por me incentivar a escrever e seguir o meu caminho. E à minha vovó, por ser uma das pessoas mais maravilhosas do mundo, por atender as minhas chamadas às 3h da manhã e por ter sido a minha companheira na altura da escrita da tese, tratando de mim com o maior carinho e dedicação.

Finally, I am grateful for every dance, every smile and every hug that replenished my soul and gave me strength to keep going…

To all of you that are my rocks!

For making me challenge myself and fly…

Abstract

The burden of infertility affects ~15% of couples worldwide with half of these cases attributed to a male factor. In order to find genetic alterations and identify disease-associated loci, researchers have been searching the genomes of infertile patients with the intent to improve diagnosis and ultimately provide new treatment options to those affected. This has proven a challenging task given the complex genetic architecture of male infertility and the broad spectrum of phenotypic outcomes. The underlying biology of sperm production and maturation to generate a viable gamete capable of fertilization is extremely complex, as discussed throughout this work, with a panoply of spermatogenesis-related genes associated with spermatogenic failure scattered throughout the genome, including the autosomes (Reviewed in paper I). In the course of this project, I adopted two different approaches to address the complex etiology of male infertility due to spermatogenic failure.

Chapter 1 describes two genetic screens performed in Portuguese patients presenting severe spermatogenic failure (SFF) but otherwise healthy and a few additional isolated cases of gonadal malformation. The candidate genes evaluated, *DMRT1* and *WT1*, were selected from a previous genome-wide association study conducted by our team and are known transcription factors with key regulatory roles in gonadal formation and function. To identify potential disease-associated variants, genetic diversity detected at these loci was assessed according to previously published guidelines for the identification of disease-causing variation. These included the estimation of variant frequency in cases versus controls and the evaluation of their potential consequences for protein sequence and function, as well as the predicted impact on transcriptional regulation and the conservation status of the affected nucleotides/residues across species. For the *DMRT1* locus (Paper II), the most promising candidates were predicted to interfere with its regulation. One novel variant in a highly conserved sequence associated with *DMRT1* repression in the promoter region (c.-223_-219CGAAA>T) was identified in one patient and predicted to disrupt the binding site for a testis-expressed heat shock protein (HSF1). Other variants with no effect on the protein sequence, two synonymous (rs376518776 and rs34946058) and two non-coding (rs144122237 and rs200423545), were enriched in the patient cohort when compared to a geographically matched control population, even though not reaching statistical significance, and an *in silico* analysis suggested that they could potentially interfere with gene expression and mRNA slicing. These results indicate that deleterious variants in *DMRT1* seem to be rare, as seen in previous studies, and may represent a risk factor for SFF through a process of gene misregulation.

Similarly, the study conducted on the *WT1* locus (Paper III) led to the identification of two rare variants overrepresented in patients and potentially relevant to disease. One novel variant (p.Pro130Leu) disrupting a mammalian-specific polyproline stretch in the self-association domain was more frequent in azoospermic patients (0.27% *versus* 0.13%, p=0.549) and a substitution (pCys350Arg) in a conserved residue in the vicinity of the first zinc-finger was more frequent in cases of severe oligozoospermia (0.80% *versus* 0.13%, p=0.113). These variants were located in exons 1 and 6, respectively, and no variation was found in the zinc-finger domain in agreement with previous studies showing that coding defects in this domain lead to syndromic phenotypes of gonadal development. The candidate-gene approach performed in chapter 1 allowed the identification of new variants potentially associated with SFF for follow-up in larger multicenter studies, suggesting that the cumulative effect of mildly deleterious variation at spermatogenesis-related loci may affect gonadal function and lead to disease phenotypes. Furthermore, these studies illustrate the challenging task of attributing causality to disease-associated genetic variation and emphasize the importance of genome annotation in terms of functional variation.

With this concept in mind, the second part of my project was dedicated to the generation of new tools that would contribute to the improvement of the functional annotation of spermatogenesis-related factors. For that, in chapter 2 I explored and developed methods to generate high-throughput data for functional genomic studies with an evolutionary perspective on spermatogenesis. The first goal was to optimize a method for isolation of specific germ-cell types from testicular samples of different mammalian species that would allow the comparative analysis of molecular events restricted to different developmental stages (Paper IV). Using fluorescence-activated cell sorting (FACS) with Hoechst-33342 DNA staining (Ho-FACS), it was possible to isolate four male germ cell populations – spermatogonia, primary/secondary spermatocytes and spermatids- from dog (*Canis familiaris*) and rat (*Rattus norvegicus*) testes with average purity of 80% estimated by microscopy. Also, an optimized gating strategy allowed the separation of round and elongating spermatid subpolulations in the mouse. This method is based on cell size, shape and complexity combined with Hoechst fluorescence and thus potentially applicable to other mammalian species as well. The work described in paper IV indicates that Ho-FACS could be used to produce highly purified subpopulations of germ cells in mammals other than mouse and that given the similar cell physiology of the mammalian spermatogenesis, it is a promising technique to be transversally applied across mammals. Importantly, the amount and integrity of the cellular material collected with this technique is sufficient and viable for high-throughput studies, as exemplified in paper V. In this study Ho-

FACS sorted cell populations were used to interrogate the translatome in different stages of spermatogenesis. Deep-sequencing of mRNA fragments undergoing translation (ribosome-protected fragments; RPFs) has been currently applied to multiple cell types from different tissues of various organisms using the recently described ribosome profiling technique. To evaluate the applicability of this method to germ cells, five relevant male germ cell populations from the mouse testis were characterized for the proportion of free mRNA, RPFs and mRNA molecules potentially repressed in ribonucleoprotein complexes. cDNA library preparation was then optimized using a commercial kit for small RNA-seq with different conditions and RNA concentrations obtained from whole mouse testis. Overall, the results indicate that ribosome profiling of Ho-FACS sorted germ cells with this optimized strategy for library preparation is not only achievable but provides a decrease in costs, time and required RNA input. Although further analysis and troubleshooting of the data generated here is required and still ongoing, it seems that differences in library preparation may have an influence on the type of transcripts detected. Importantly, ribo-seq data from cDNA libraries prepared using 10µg or 100 ng of RPFs indicates that using reduced amounts of RNA input does not reduce the depth of sequenced transcript diversity ($r^2 = 0.81$).

Overall, the work described here provides a comprehensive genetic characterization of two loci associated with phenotypes of gonadal dysfunction that supports a view where relevant mildly deleterious rare variants are scattered throughout the genome and represents risk factors that collectively contribute to the phenotype. Also, it provides novel tools for functional genomic studies that hold the promise to bring new insights into the regulatory networks governing spermatogenesis. This information is crucial for the annotation of functionally relevant spermatogenesis-related genes and can ultimately aid in the assessment of the functional impact of variation identified by genome-wide studies on patient cohorts. The integration of such knowledge and translation to the clinical setting can open new avenues in the diagnosis, treatment and general counselling of male infertile patients.

Key words: Male infertility, DMRT1, WT1, FACS, Ribosome Profiling.

Sumário

Globalmente, a infertilidade afeta cerca de 15% dos casais sendo que em metade dos casos é devida a causas masculinas. De forma a identificar alterações genéticas em loci associados à doença, os investigadores têm feito pesquisas no genoma de doentes inférteis com o objetivo de melhorar os métodos de diagnóstico e desenvolver novas opções de tratamento. Esta tarefa tem sido um desafio devido à complexa arquitecura genética da infertilidade masculina e o largo espectro de fenótipos que esta apresenta. A base biológica da produção e maturação do esperma para gerar um gâmeta viável e com capacidade de fertilização é extremamente complexa, conforme discutido ao longo deste trabalho, com uma panóplia de genes envolvidos na espermatogénese espalhados pelo genoma, incluindo nos autosomas (Revisto no artigo I), associados a insuficiência da espermatogénese. Neste projeto adotei duas estratégias diferentes para abordar a etiologia da infertilidade masculina causada por insuficiência da espermatogénese.

O capítulo 1 descreve dois estudos genéticos efetuados em doentes Portugueses com insuficiência severa da espermatogénese ("severe spermatogenic failure" - SFF) mas saudáveis, e alguns casos pontuais de doentes com malformação das gónadas. Os genes candidatos avaliados, *DMRT1* e *WT1*, foram selecionados no seguimento de um estudo de associação abrangendo todo o genoma, realizado previamente pela nossa equipa e são fatores de transcrição reconhecidos pelos seus papéis reguladores essenciais para a formação e manutenção da função das gónadas. A fim de identificar potenciais variantes associados à doença, a diversidade genética detetada nestes loci foi aferida tendo em conta as normas publicadas para a identificação de variantes causadores de doença. Estas incluíram a estimativa da frequência dos variantes encontrados em doentes versus controlos e a avaliação das suas potenciais consequências ao nível da sequência e função da proteína, bem como o impacto previsto na regulação da transcrição e a conservação dos nucleótidos/resíduos afetados entre diferentes espécies. No estudo do gene *DMRT1* (artigo II), as previsões *in silico* apontam para que os variantes mais promissores possam interferir com a sua regulação. Numa região altamente conservada do promotor, previamente associada à repressão do gene *DMRT1*, foi identificado um variante (c.-223_-219CGAAA>T) que se prevê interferir a ligação de uma proteína de choque térmico (HSP1) expressa em testículo. Outros variantes sem consequências para a sequência da proteína, dois sinónimos (rs376518776 and rs34946058) e dois não codificantes (rs144122237 and rs200423545), estavam sobrerepresentados em doentes quando comparados com uma população controlo da mesma região geográfica e uma

análise *in silico* sugeriu que estes podem potencialmente interferir com a expressão do gene e processamento do transcrito. Estes resultados indicam que variantes deletérios no gene *DMRT1* serão raros, conforme reportado em estudos anteriores, e poderão representar fatores de risco para SFF através de um processo de desregulação do gene. Da mesma forma, o estudo realizado no gene *WT1* (artigo III) permitiu a identificação de dois variantes raros, sobrerepresentados em doentes, e potenticalmente relevantes para a doença. Um variante novo no domínio de homodimerização (p.Pro130Leu) que quebra uma sequência de poliprolinas específica de mamíferos, é mais frequente em casos de azoospermia (0.27% *versus* 0.13%, p=0.549), enquanto que a substituição (pCys350Arg) num resíduo conservado próximo do primeiro zinc-finger da proteína é mais frequente em casos de oligozoospermia severa (0.80% *versus* 0.13%, p=0.113). Estes variantes localizam-se nos exões 1 e 6, respetivamente, e a ausência de variação no domínio zinc-finger está em concordância com estudos anteriores demonstrando que alterações codificantes neste domínio resultam em síndromes em que se manifestam anomalias no desenvolvimento das gónadas. O estudo de genes candidatos descrito no capítulo 1 permitiu a identificação de novos variantes possivelmente associados com SFF que deverão ser validados em estudos mais abrangentes e sugerem que um efeito cumulativo de variantes moderadamente deletérios em loci envolvidos na espermatogénese pode afetar a função das gónadas e resultar em fenótipos de infertilidade. Ademais, estes estudos demonstram a dificuldade em definir variantes como causa de doença, reforçando a importância de anotar o genoma em termos de variação funcional.

Seguindo este conceito, a segunda parte do meu projeto foi dedicada ao desenvolvimento de novos métodos que contribuíssem para o melhoramento da anotação funcional de fatores envolvidos na espermatogénese. Para tal, no capítulo 2 explorei e desenvolvi novos protocolos para gerar dados com elevado rendimento em estudos de genómica funcional com uma perspetiva evolutiva da espermatogénese. O primeiro objetivo consistiu em otimizar um método para o isolamento de diferentes tipos de células germinais a partir de amostras testiculares, que possibilitasse uma análise comparativa dos eventos moleculares específicos de diferentes estadios do desenvolvimento (artigo IV). Através da técnica de separação de células ativada por fluorescência (“fluorescence-activated cell sorting” - FACS) do marcador de ADN Hoechst-33342 (Ho-FACS) foi possível isolar quatro populações de células germinais – espermatogónias, espermatócitos primários/secundários e espermatídeos – de testículo canino (*Canis familiaris*) e de ratazana (*Rattus norvegicus*) com uma homogeneidade média de 80% estimada por microscopia. Adicionalmente, foi otimizada

no ratinho uma estratégia para distinguir subpopulações de espermatídeos redondos e em alongamento. Como este método se baseia numa combinação entre a fluorescência do marcador Hoechst e características celulares de tamanho, forma e complexidade é potencialmente aplicável a outras espécies de mamíferos. O trabalho descrito no artigo IV indica que a técnica de Ho-FACS pode ser usada para purificar subpopulações de células germinais noutras espécies de mamíferos que não o ratinho e que, tendo em conta a semelhança na fisiologia celular da espermatógenese em mamíferos, é uma técnica promissora para uma aplicação transversal dentro dos organismos deste grupo. É de realçar que a quantidade e a integridade do material isolado com este técnica é suficiente e viável para estudos de elevado rendimento, conforme demonstrado no artigo V. Neste estudo, populações de células germinais isoladas por Ho-FACS foram utilizadas para investigar o translatoma em diferentes estadios da espermatogénese. A sequenciação de fragmentos de mARN em tradução ("ribosome-protected fragments" – RPFs) tem sido aplicada a vários tipos de células de diferentes tecidos de diversos organismos através da técnica recentemente descrita de perfil dos fragmentos de mARN protegidos pelo ribossoma ("Ribosome Profiling"). Com o intuito de avaliar a aplicabilidade deste método a células germinais, foram caracterizadas cinco populações de células germinais masculinas em termos de proporção de mARN livre, FPRs, e moleculas de mARN potencialmente reprimidas em complexos ribonucleoproteicos. De seguida, foi otimizada a preparação de bibliotecas de cADN recorrendo a um kit comercial desenvolvido para sequenciação de ANRs de pequeno tamanho fazendo variar diferentes condições técnicas e concentrações de ARN extraído de testículo total de ratinho. Os resultados indicam que a aplicação da técnica de "Ribosome Profiling" a células germinais isoladas por Ho-FACS não só é possível como ainda proporciona uma redução nos custos, tempo e quantidade de ARN necessário. A análise dos dados gerados neste trabalho ainda está a decorrer mas os dados preliminares sugerem que diferentes métodos de preparação de bibliotecas podem influenciar o tipo de transcritos detetados por sequenciação. Contudo, os dados obtidos de bibliotecas de cADN preparadas a partir de 10 µg ou 100 ng de RPFs indicam que a redução da quantidade de ARN não compromete a diversidade de transcritos sequenciados.

Em suma, o trabalho descrito aqui proporciona uma caracterização genética compreensiva de dois *loci* associados a fenótipos de disfunção das gónadas que suporta a hipótese de que variantes deletérios de impacto mais moderado estão distribuídos pelo genoma representando fatores de suscetibilidade que coletivamente contribuem para o desenvolvimento do fenótipo. Adicionalmente, providencia novas ferramentas para estudos de

genómica funcional que poderão trazer novo conhecimento sobrea rede de regulação que controla a espermatogénese. Esta informação é crucial para a anotação de genes funcionalmente relevantes para a espermatogénese, contribuindo para a avaliação do impacto funcional de variantes identificados através de estudos de genómica em coortes de doentes. A integração deste conhecimento e a sua translação para a clínica poderão abrir novas possibilidades para o diagnóstico, tratamento e aconselhamento de doentes com infertilidade masculina.

Palavras-chave: Infertilidade masculina, DMRT1, WT1, FACS, Ribosome Profiling.

Table of Contents

Abbreviations

2D — Two-dimensional

> — greater than

< — smaller than

~ — Approximatelly

A

AMH — Anti-Müllerian Hormone

ARTs — Assisted reproductive techniques

B

Bp — Base pair

BSA — Bovine serum albumin

BTB — Blood–testis barrier

C

cDNA — Complementary DNA

CDS — Coding sequence

CHX — Cycloheximide

CNV — Copy-number variant

D

DNA — Deoxyribonucleic Acid

DMRT1 — Doublesex and mab-3 related transcription factor 1

E

ENCODE — Encyclopedia of DNA Elements project

eSpd — elongating spermatid

ESC — Embryonic stem cell

EtOH — Ethanol

F

FACS — Fluorescence-activated cell sorting

FANTOM — Functional Annotation of the Mammalian Genome project

FOXL2 — Forkhead box L2

FS — Forward scatter

FSH — Follicle stimulating hormone

G

GC — Germ cell

GCDmrt1— Dmrt1 produced by germ cells

GC*Dmrt1*KO — Conditional KOs of *Dmrt1* in germ cells

GD — Gonadal dysgenesis

GnRH — Gonadotropin-releasing hormone

GO — Gene ontology

GTEx — Genotype-Tissue Expression project

GWAS — Genome-wide association study

H

HI — Haploinsufficient

Ho — Hoechst-33342

Ho-FACS — Fluorescence-activated cell sorting with Hoechst-33342 staining

HPG-axis — Hypothalamic-pituitary-gonadal axis

I

ICSI — Intracytoplasmic sperm injection

IFC — Imaging flow cytometry (IFC)

Indel — Insertion/deletion

K

Kb — Kilo-base pair

KO — Knock-out

KTS — Lysine, threonine and serine

L

LH — Luteinizing hormone

lncRNA — Long non-coding RNA

LoF — Loss-of-function

M

MACS — Magnetic-activated cell sorting

min-- Minute

miRNA — Micro-RNA

mRNA — Messenger ribonucleic acid

N

NOA — Non-obstructive Azoospermia

NR5A1 — Nuclear receptor subfamily 5, group A, member 1

nt— nucleotide

P

PAGE — Polyacrylamide gel electrophoresis

PCR — Polymerase chain reaction

PGCLCs — Embryonic stem cell-derived primordial germ cells

PI — Propidium Iodine

piRNA— Piwi-interacting RNA

PIT — Proteomics Informed by Transcriptomics

preL— pre-leptotene spermatocyte

R

r^2 — correlation coefficient

RA — Retinoic Acid

RBP — RNA-binding protein

Ribo-seq — deep-sequencing of ribosome-protected mRNA fragments

RNA — Ribonucleic Acid

rRNA — Ribosomal RNA

RNP — Ribonucleoproteins

RPF — Ribosome-protected mRNA fragment

rSpd — round spermatid

RT— Room temperature

RT-PCR — Reverse transcription polymerase chain reaction

S

s — Second

SC — Sertoli cell

SC*Dmrt1*KO — Conditional KOs of *Dmrt1* in Sertoli cells

SEC — Size exclusion chromatography

SF1 — Steroidogenic Factor 1

sncRNAs — Small non-coding RNAs

SNP — Single nucleotide polymorphism

SNV — Single nucleotide variant

SOX9 — SRY box 9

SOZ — Severe oligozoospermia

Spg — Spermatogonia

Spc I — Primary spermatocyte

Spc II — Secundary spermatocyte

Spd — Spermatid

SRY — Sex-determining region Y

SS — Side scatter

SSF — Severe spermatogenic failure

STRA8 — Stimulated By Retinoic Acid 8

T

TE — Transposable elements

TFBS — Transcription factor binding-site

TIS — Translation Initiation Site

TSD — Temperature-dependent sex determination

TSS — Transcription start site

TUT — Testicular unannotated transcripts

U

U— Units

uORFs — Upstream open reading frames

UTR — Untranslated region

V

VDG — Vybrant® DyeCycle™ Green Stain

W

WNT — Wingless-type MMTV (mouse mammary tumor virus) integration site

WT — Wild-type

WT1 — Wilms' Tumor 1

Z

ZF — Zinc-finger

List of Figures

General Introduction

Paper I

Chapter 1

Introduction

Paper II

Paper III

Chapter 2

Introduction

Paper IV

Paper V

Appendix

Paper II

List of Tables

General Introduction

Paper I

Chapter 1

Specific objectives & Results

Paper II

Paper III

Chapter 2

Introduction

Appendix

Paper II

General Introduction

The development of highly specialized reproductive mechanisms throughout evolution has likely been fine-tuned by sexual selection, due to individual advantage within a species (Darwin 1871). One of such was sexual dimorphism, represented in its early forms by anisogamy, the differentiation between male and female gametes (Lehtonen *et al.* 2016). This is thought to have resulted from gamete competition and selection in complex multicellular organisms, where males typically produce larger numbers of smaller gametes when compared to females (Lehtonen and Parker 2014; Parker and Lehtonen 2014). In diploid species, the process of male gamete formation - spermatogenesis - essentially ensures the development of haploid cells capable of fertilization and transmission of the genetic patrimony. The basic mechanisms of spermatogenesis are shared by many different organisms and the genes involved in these processes are highly conserved among mammals (White-Cooper and Bausek 2010). Alterations in genes involved in the network controlling spermatogenesis result in disease with phenotypes of varying severity, ranging from milder alterations in sperm parameters to syndromes that compromise gonadal development (For reviews see Matzuk and Lamb (2002); Ferlin *et al.* (2006); Carrel (2007); Matzuk and Lamb (2008); Hwang *et al.* (2010); Massart *et al.* (2012); Paper I).

An overview of mammalian spermatogenesis

Spermatogenesis takes place within the seminiferous tubules of the testis. They are supported by the testicular interstitium that provides blood supply, immunological responses and mediates endocrine signals to and from the pituitary through the Leydig cells. The external layer of the seminiferous tubules contains myoid peritubular cells that produce the internal basal membrane, which encloses the germ line and the somatic Sertoli cells (Reviewed in Wistuba *et al.* (2007)). Many traits can present species-specificity, such as the topographic arrangement of testicular cells in different stages of the seminiferous epithelium cycle and Sertoli cell abundance, which ultimately determines the number of spermatozoa produced and testicular size in mammals (Reviewed in Wistuba *et al.* (2007)). Interestingly, the basic mechanisms of spermatogenesis progression are shared across mammals (White-Cooper and Bausek 2010) and comprise the mitotic, meiotic and spermiogenesis phases (Hess and Renato de Franca 2008). Figure 1 illustrates the cellular organization of the human seminiferous epithelium and the main developmental stages throughout spermatogenesis progression directed from the periphery towards the lumen of the tubule. At the basal membrane, spermatogonial stem cells undergo several rounds of mitosis that both replenish the stem cell pool and generate

proliferating spermatogonia that will later commit to differentiation. Four classes of spermatogonia have been described in rodents (Reviewed in Hess and Renato de Franca (2008)): undifferentiated type A spermatogonia [A single (A_s), A paired (A_{pr}), A aligned (A_{al})]; differentiated type A spermatogonia (A1, A2, A3, A4); intermediate spermatogonia (In); and type B spermatogonia (B)], however there is still a debate regarding which populations represent the true stem cell pool (Mei *et al.* 2015). B-spermatogonia divide in two pre-Leptotene (preL) spermatocytes that initiate meiosis S-phase with DNA replication originating cells with 4N DNA content (Baarends and Grootegoed 2003). Chromatin configuration and structure further define 5 Prophase I substages: Leptotene, Zygotene, Pachytene, Diplotene and Diakinesis, which are accompanied by the major events of chromosome synapsis and homologous recombination (Reviewed in Baarends and Grootegoed (2003); Handel and Schimenti (2010)). Then, the 2N secondary spermatocytes undergo the second round of meiosis to generate haploid round spermatids that enter spermiogenesis. During this phase, both cell shape and chromatin structure change drastically, progressively elongating and condensing, respectively, while histones are replaced by protamines. This special chromatin arrangement is thought to help generating a compact hydrodynamic shape, protect the paternal genome from damage and allow for specific epigenetic regulation (Rathke *et al.* 2014). When this process is completed, the contractile myoid peritubular cells direct the immotile testicular spermatozoa from the Sertoli cells to the efferent ducts. From there, spermatozoa enter the epididymis, a long tubular structure with a specialized milieu that promotes sperm maturation. Once they pass the epididymis, spermatozoa are motile and have acquired a series of modifications that are required for fertilization (Reviewed in Dacheux and Dacheux (2014). The last step of sperm maturation - capacitation - occurs within the female reproductive tract where spermatozoa become competent for fertilization (Reviewed in Aitken and Nixon (2013)).

Gonadal structure and function are regulated by a highly orchestrated interplay of many molecular factors, where the endocrine regulation established by the hypothalamic-pituitary-gonadal (HPG) axis plays a central role. Briefly, the hypothalamus releases gonadotropin-releasing hormone (GnRH) that stimulates the pituitary gland to produce the gonadotropins luteinizing hormone (LH) and follicle stimulating hormone (FSH). Leydig and Sertoli cells in the testis are sensitive to LH and FSH, respectively, and respond with the production of androgens and Inhibin. In return, circulating androgens and Inhibin trigger the production of LH and FSH maintaining a positive feedback loop. The FSH-inhibin and LH-testosterone feedback loops are two fully independent mechanisms that act specifically in the regulation of spermatogenesis and steroidogenesis, respectively. The former is mediated by Sertoli cells in the testis and

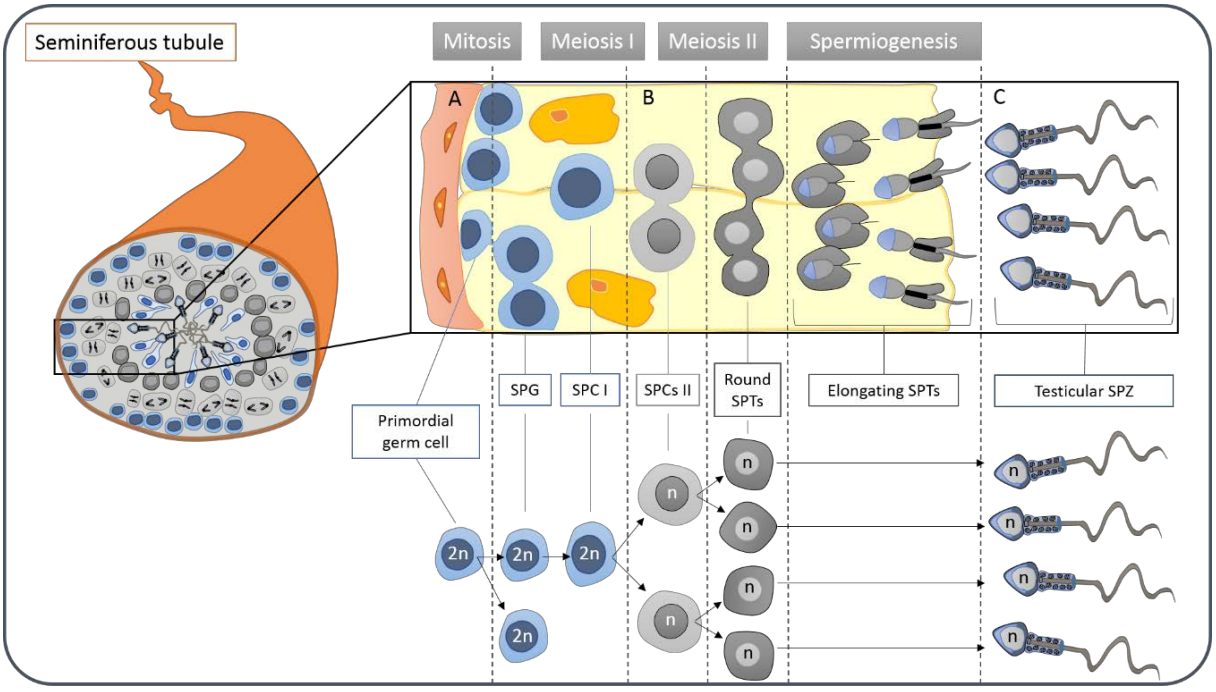


Figure 1. Overview of human spermatogenesis.

In this schematics the process of spermatogenesis is depicted in an overview of a cross-section of a seminiferous tubule. Male gametes develop within the epithelia (B- Sertoli Cells) of the seminiferous tubules in the testis, with differentiation occurring from the wall (A) towards the lumen (C). Diploid germ cells undergo several rounds of mitosis giving rise to type A spermatogonia (to maintain the pool of stem cells) and type B spermatogonia (SPG) that differentiate into primary spermatocytes (SPC I). These cells then go through meiosis I, where recombination of genetic material occurs, and originate haploid secondary spermatocytes (SPCs II). A second round of meiosis originates haploid round spermatids (SPTs) that, through spermiogenesis, acquire sperm cell specializations and are released (spermiation) into the lumen, becoming testicular spermatozoa (SPZ). *In* Lima and Lopes (2014) - Paper I.

controls the expansion of premeiotic germ cells. Testosterone produced by Leydig cells upon LH stimulation is responsible for maintaining the male phenotype (Reviewed in Schlatt and Ehmcke (2014)). For the scope of this work I will focus on the molecular intracellular mechanisms playing a role in spermatogenesis progression. Specifically, I will address the impact of genetic variants in crucial regulators of this developmental process and the regulatory layers acting on the different processing steps of protein production from gene expression to translation.

Perhaps one of the most surprising observations brought by genome technologies was that the number of estimated genes in mammalian genomes (~22,000; Hubbard *et al.* (2007)) is far more reduced than expected, given the high level of catalogued protein diversity

(Humphery-Smith 2004; Mueller *et al.* 2007). This is known to result from, and can only be explained by the complexity of multilayered mechanisms of gene expression regulation reported in mammalian cells. Actually, the introduction of elaborate gene regulatory circuits is thought to have been one of the driving forces of eukaryotic evolution, given the apparent lack of correlation between genome size and organismal complexity (Gregory 2001). Figure 2 depicts the main regulatory layers within eukaryotic cells. In this concern, germ cells appear to have evolved a very elaborate regulatory network. For instance, many genes are specifically or differentially expressed in the germ cells (Chalmel *et al.* 2007) with complex mechanisms of alternative splicing (Elliott and Grellscheid 2006). Also, long non-coding (lncRNAs) and small non-coding (sncRNAs) RNAs, such as microRNAs (miRNAs) and Piwi-interacting RNAs (piRNAs) are major players in posttranscriptional regulation and essential for normal spermatogenesis progression (For reviews see de Mateo and Sassone-Corsi (2014); Luk *et al.* (2014)). These non-coding RNAs, as well as the RNA-binding proteins (RBPs), are known to be involved in translational repression of mRNAs, which is a common mechanism of transcript regulation extensively used by germ cells (Iguchi *et al.* 2006; Bettegowda and Wilkinson 2010; Gan *et al.* 2013; de Mateo and Sassone-Corsi 2014). At the protein level, as an example, both proteolytic and non-proteolytic functions of the ubiquitin-protease system are crucial during spermatogenesis, being required for proper spermatogonial development, meiosis, meiotic sex chromosome inactivation and spermiogenesis (Reviewed in Bose *et al.* (2014)). Integrating knowledge of the mechanisms involved in all of these different regulatory layers has the potential to elucidate the molecular interactions driving spermatogenesis but also to identify candidate genes for infertility and for the development of diagnostic tools and contraceptives.

Male infertility and the efforts to tackle the genetic etiology of complex phenotypes

Infertility affects approximately 15% of couples worldwide, with half the cases attributed to a male factor. A diagnosis can be established in 60%-70% of the cases and the remaining are classified as idiopathic male infertility (Jungwirth *et al.* 2013). Considering the intricate molecular biology of the male reproductive system, regulated by a panoply of factors, it is reasonable to admit that disturbances in this delicate balance can interfere with different aspects of gonadal development or function and lead to variable phenotypic outcomes. Severe spermatogenic failure (SFF) is manifested by a drastic reduction in sperm numbers that can be accompanied by alteration of other sperm parameters (motility and/or cell morphology) or, ultimately, absence of sperm from the ejaculate - azoospermia ((Practice Committee of

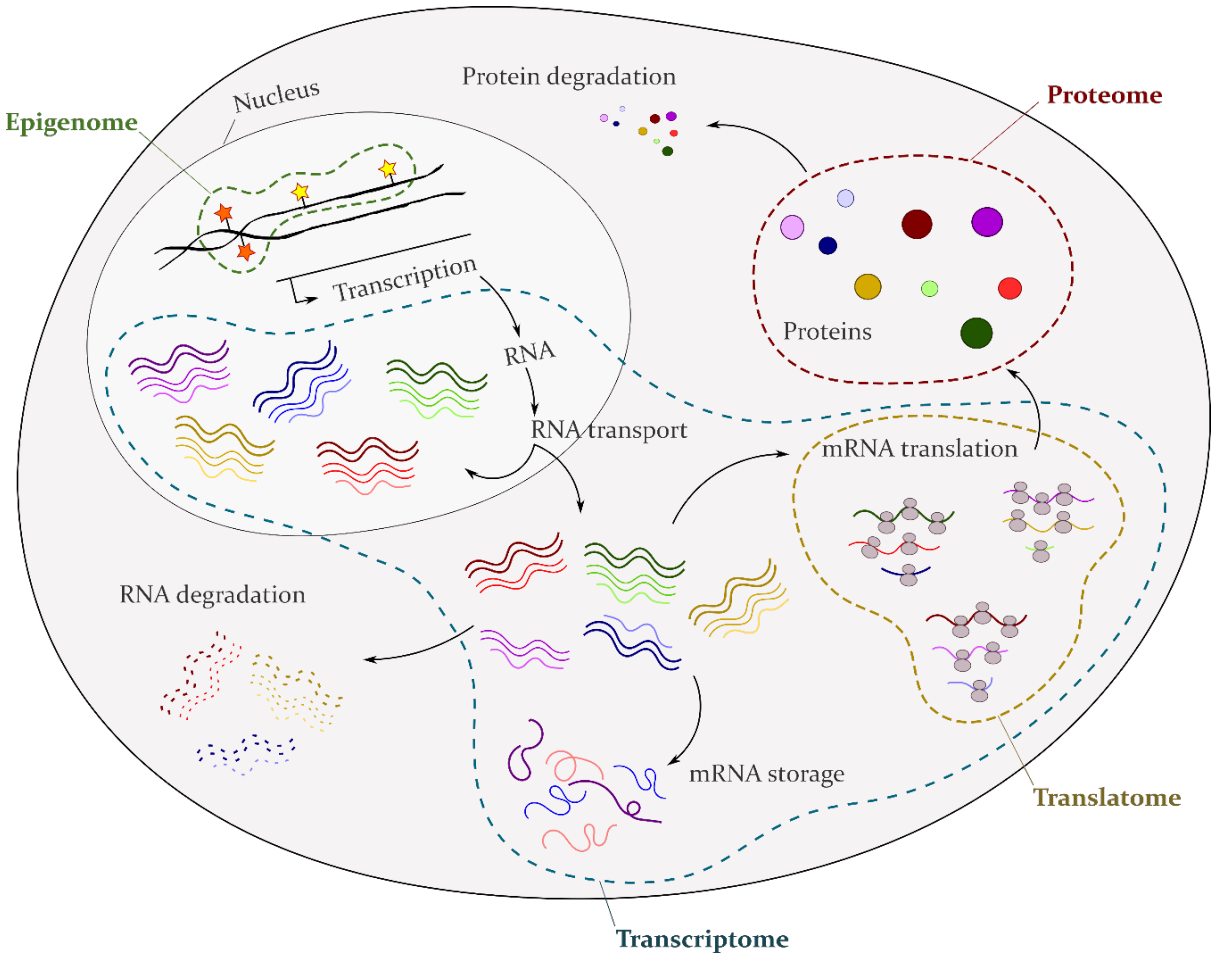


Figure 2. Regulatory mechanisms of transcript production and processing in mammalian cells.
Essentially, gene expression is mediated by epigenetic modifications of chromatin (epigenome) and promoter-binding elements such as transcription factors, inter- and intragenic cis-regulatory elements (enhancers/silencers) and non-coding RNAs (regulome). Transcribed mRNAs (transcriptome) can originate many transcript isoforms (spliceome) by means of alternative splicing, which in turn can be regulated at the level of transport, storage and translation. Moreover, transcripts undergoing translation (translatome) can influence the outcome of protein synthesis by the recruitment of specific ribosomal RNAs. Finally, posttranslational modifications (e.g., phosphorylation, glycosylation and proteolysis) will influence protein function (proteome). This complex mechanism controls homeostasis of any given somatic cell and is behind the production of tissue-specific proteins. Adapted from (Piccirillo et al. 2014).

American Society for Reproductive Medicine in collaboration with Society for Male Reproduction and Urology 2008; Jungwirth *et al.* 2013); Paper I). Importantly, the ethiopathogenesis of SFF in otherwise healthy men (non-syndromic form) is complex and still largely unknown. The work in Paper I reviews relevant studies that have reported an association of different autosomal loci with specific phenotypes of SSF. Most of these works

have followed a candidate-gene strategy and have contributed greatly to improve our knowledge of genetic defects involved in this multifactorial phenotype. However, advances in whole-genome sequencing technologies have brought the exciting possibility of collecting information from multiple relevant loci and identify new genes in patients with SSF. Also, the continuous drop in sequencing costs per human genome allows for studies with larger patient cohorts (Carrell *et al.* 2016), which is critical for complex diseases associated with rare variants (Tennessen *et al.* 2012). Yet, the application of genomics to health relies on prior knowledge of gene pathways playing a role in disease and how they interact with environmental factors, going beyond the identification of fertility-related genes towards the understanding of the relevance of regulatory molecules, such as micro-RNA, enhancers, promoters, and others, in disease development (Carrel 2007). In this regard, functional genomics is the field dedicated to the elucidation of the molecular basis of biological functions making use of next-generation sequencing technologies (Morozova and Marra 2008; Werner 2010).

By analogy with the word “genome”, which refers to all the genetic material within a given cell, pools of different types of molecules sharing a common trait have been defined (Fig. 2). For instance, transcriptome and proteome represent total RNA or protein content within a biological system, respectively, whereas translatome denotes all transcripts undergoing translation at a given time. Similarly, the tools and techniques applied to the global analysis of such groups of molecules are named in respect to the biological material analyzed with the addition of the suffix -*omics*. For instance, transcriptomics and proteomics comprise the methods employed in the collection of transcriptome and proteome data, respectively. Detailed descriptions of the different methods, advantages and drawbacks can be found elsewhere (Roy *et al.* 2011; Haider and Pal 2013; Mooney and McWeeney 2014; Boersema *et al.* 2015). In general, the most significant improvement brought by these techniques was the possibility to study simultaneously tens of thousands of gene products rather than working on a gene-by-gene basis. This technological revolution has prompted the generation of genotype and functional data from many different tissues, especially for human and mouse, by ever-larger scale projects such as the Encyclopedia of DNA Elements (ENCODE; Encode Project Consortium (2012)), Genotype-Tissue Expression Project (GTEx; G. TEx Consortium (2013)), Human Protein Atlas (Uhlen *et al.* 2010; Uhlen *et al.* 2015), and Functional Annotation of the Mammalian Genome (FANTOM; Kawai *et al.* (2001)). Despite the inherent challenge of identifying the true biological signal within such large and complex datasets, these data hold the promise of uncovering gene function, regulation, and even contributing to the understanding of the genetic basis of complex diseases (Greene and Troyanskaya 2012). Indeed, one great

example of the potential of functional genomics to identify fertility-related genes was already demonstrated by Guan and co-workers (Guan *et al.* 2012). The authors developed and algorithm that integrates diverse functional genomics data to generate tissue-specific functional networks in the mouse and show an improved accuracy in predicting phenotype-related genes when compared to a single global functional network. Furthermore, the authors used this approach to predict male infertility and spermatogenesis-related genes, using a testis-specific functional relationship network, and the top hit was experimentally validated. More recently, Zhu and co-authors investigated the dynamics of human spermatogenesis and identified differently expressed genes that can potentially serve as molecular tools for clinical purposes (Zhu *et al.* 2016). Additionally, the authors reported 157 differently expressed lncRNAs, which suggests a relevant biological role for these molecules during spermatogenesis progression and further emphasizes the importance of functional genomics to tackle the complex genetic etiology of SFF. Currently, the standard genetic tests only allow the diagnosis of ~20% of male infertility cases and new promising diagnostic tools include epigenetic analysis of spermatozoa and detection of rare genetic variants and copy-number variation (Hotaling and Carrell 2014). Functional data has been demonstrating its power in identifying novel causes of male infertility (Carrell *et al.* 2016) and the effort to integrate these diverse datasets should continue to bring novel tools for the diagnosis and treatment of male infertility.

The ultimate goal of this work was to investigate the regulatory networks of spermatogenesis and the impact of genetic variants in regulatory factors on male infertility. My approach combines two different strategies:

1) Search for deleterious variants in two key regulators of spermatogenesis in men with SSF, using a candidate gene approach in case-control studies (Chapter 1);

2) Develop new tools in a mouse model to study the transcriptome and the translatome in different stages of spermatogenesis in order to identify relevant mammalian spermatogenesis regulators and gain insights into the functional relevance of regulatory and coding genetic variants found in the context of male infertility (Chapter 2).

Literature Cited

Aitken, R.J. and Nixon, B. (2013). Sperm capacitation: a distant landscape glimpsed but unexplored. *Mol Hum Reprod,* **19**: 12, 785-793.

Baarends, W.M. and Grootegoed, J.A. (2003). Chromatin dynamics in the male meiotic prophase. *Cytogenet Genome Res,* **103**: 3-4, 225-234.

Bettegowda, A. and Wilkinson, M.F. (2010). Transcription and post-transcriptional regulation of spermatogenesis. *Philos Trans R Soc Lond B Biol Sci,* **365**: 1546, 1637-1651.

Boersema, P.J., Kahraman, A. and Picotti, P. (2015). Proteomics beyond large-scale protein expression analysis. *Curr Opin Biotechnol,* **34**: 162-170.

Bose, R., Manku, G., Culty, M. and Wing, S.S. (2014). Ubiquitin-proteasome system in spermatogenesis. *Adv Exp Med Biol,* **759**: 181-213.

Carrel, D.T. (2007). The Genetics of Male Infertility in the Era of Genomics: Tools for Progress *in*: The Genetics of Male Infertility, Humana Press Inc., I, 1-58829-863-9.

Carrell, D.T., Aston, K.I., Oliva, R., Emery, B.R. and De Jonge, C.J. (2016). The "omics" of human male infertility: integrating big data in a systems biology approach. *Cell Tissue Res,* **363**: 1, 295-312.

Chalmel, F., Lardenois, A. and Primig, M. (2007). Toward understanding the core meiotic transcriptome in mammals and its implications for somatic cancer. *Ann N Y Acad Sci,* **1120**: 1-15.

Consortium, E.P. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature,* **489**: 7414, 57-74.

Consortium, G.T. (2013). The Genotype-Tissue Expression (GTEx) project. *Nat Genet,* **45**: 6, 580-585.

Dacheux, J.L. and Dacheux, F. (2014). New insights into epididymal function in relation to sperm maturation. *Reproduction,* **147**: 2, R27-42.

Darwin, C. (1871). Sexual Selection *in*: The descent of man and selection in relation to sex, 1, Princeton University Press, II, 0-691-08278-2.

de Mateo, S. and Sassone-Corsi, P. (2014). Regulation of spermatogenesis by small non-coding RNAs: role of the germ granule. *Semin Cell Dev Biol,* **29**: 84-92.

Elliott, D.J. and Grellscheid, S.N. (2006). Alternative RNA splicing regulation in the testis. *Reproduction,* **132**: 6, 811-819.

Ferlin, A., Arredi, B. and Foresta, C. (2006). Genetic causes of male infertility. *Reprod Toxicol,* **22**: 2, 133-141.

Gan, H., Cai, T., Lin, X., Wu, Y., Wang, X., Yang, F. and Han, C. (2013). Integrative proteomic and transcriptomic analyses reveal multiple post-transcriptional regulatory mechanisms of mouse spermatogenesis. *Mol Cell Proteomics,* **12**: 5, 1144-1157.

Greene, C.S. and Troyanskaya, O.G. (2012). Accurate evaluation and analysis of functional genomics data and methods. *Ann N Y Acad Sci,* **1260**: 95-100.

Gregory, T.R. (2001). Coincidence, coevolution, or causation? DNA content, cell size, and the C-value enigma. *Biol Rev Camb Philos Soc,* **76**: 1, 65-101.

Guan, Y., Gorenshteyn, D., Burmeister, M., Wong, A.K., Schimenti, J.C., Handel, M.A., Bult, C.J., Hibbs, M.A. and Troyanskaya, O.G. (2012). Tissue-specific functional networks for prioritizing phenotype and disease genes. *PLoS Comput Biol,* **8**: 9, e1002694.

Haider, S. and Pal, R. (2013). Integrated analysis of transcriptomic and proteomic data. *Curr Genomics,* **14**: 2, 91-110.

Handel, M.A. and Schimenti, J.C. (2010). Genetics of mammalian meiosis: regulation, dynamics and impact on fertility. *Nat Rev Genet,* **11**: 2, 124-136.

Hess, R.A. and Renato de Franca, L. (2008). Spermatogenesis and cycle of the seminiferous epithelium. *Adv Exp Med Biol,* **636**: 1-15.

Hotaling, J. and Carrell, D.T. (2014). Clinical genetic testing for male factor infertility: current applications and future directions. *Andrology,* **2**: 3, 339-350.

Hubbard, T.J., Aken, B.L., Beal, K., Ballester, B., Caccamo, M.*, et al.* (2007). Ensembl 2007. *Nucleic Acids Res,* **35**: Database issue, D610-617.

Humphery-Smith, I. (2004). A human proteome project with a beginning and an end. *Proteomics,* **4**: 9, 2519-2521.

Hwang, K., Yatsenko, A.N., Jorgez, C.J., Mukherjee, S., Nalam, R.L., Matzuk, M.M. and Lamb, D.J. (2010). Mendelian genetics of male infertility. *Ann N Y Acad Sci,* **1214**: E1-E17.

Iguchi, N., Tobias, J.W. and Hecht, N.B. (2006). Expression profiling reveals meiotic male germ cell mRNAs that are translationally up- and down-regulated. *Proc Natl Acad Sci U S A,* **103**: 20, 7712-7717.

Jungwirth, A., Diemer, T., Dohle, G., Giwercman, A., Kopa, Z., Krausz, C. and Tournaye, H. (2013). Guidelines on Male Infertility *in*: 12-05-2014, Uroweb,

Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M.*, et al.* (2001). Functional annotation of a full-length mouse cDNA collection. *Nature,* **409**: 6821, 685-690.

Lehtonen, J. and Parker, G.A. (2014). Gamete competition, gamete limitation, and the evolution of the two sexes. *Mol Hum Reprod,* **20**: 12, 1161-1168.

Lehtonen, J., Parker, G.A. and Scharer, L. (2016). Why anisogamy drives ancestral sex roles. *Evolution,*

Lima, A.C. and Lopes, A.M. (2014). Autosomal Mutations and Spermatogenic Failure *in*: Encyclopedia of Life Sciences, eLS. John Wiley & Sons, Ltd,

Luk, A.C., Chan, W.Y., Rennert, O.M. and Lee, T.L. (2014). Long noncoding RNAs in spermatogenesis: insights from recent high-throughput transcriptome studies. *Reproduction,* **147**: 5, R131-141.

Massart, A., Lissens, W., Tournaye, H. and Stouffs, K. (2012). Genetic causes of spermatogenic failure. *Asian J Androl,* **14**: 1, 40-48.

Matzuk, M.M. and Lamb, D.J. (2002). Genetic dissection of mammalian fertility pathways. *Nat Cell Biol,* **4 Suppl**: s41-49.

Matzuk, M.M. and Lamb, D.J. (2008). The biology of infertility: research advances and clinical challenges. *Nature Medicine,* **14**: 11, 1197-1213.

Mei, X.X., Wang, J. and Wu, J. (2015). Extrinsic and intrinsic factors controlling spermatogonial stem cell self-renewal and differentiation. *Asian J Androl,* **17**: 3, 347-354.

Mooney, M. and McWeeney, S. (2014). Data integration and reproducibility for high-throughput transcriptomics. *Int Rev Neurobiol,* **116**: 55-71.

Morozova, O. and Marra, M.A. (2008). Applications of next-generation sequencing technologies in functional genomics. *Genomics,* **92**: 5, 255-264.

Mueller, M., Martens, L. and Apweiler, R. (2007). Annotating the human proteome: beyond establishing a parts list. *Biochim Biophys Acta,* **1774**: 2, 175-191.

Parker, G.A. and Lehtonen, J. (2014). Gamete evolution and sperm numbers: sperm competition versus sperm limitation. *Proc Biol Sci,* **281**: 1791, 20140836.

Practice Committee of American Society for Reproductive Medicine in collaboration with Society for Male Reproduction and Urology (2008). Evaluation of the azoospermic male. *Fertil Steril,* **90**: 5 Suppl, S74-77.

Rathke, C., Baarends, W.M., Awe, S. and Renkawitz-Pohl, R. (2014). Chromatin dynamics during spermiogenesis. *Biochim Biophys Acta,* **1839**: 3, 155-168.

Roy, N.C., Altermann, E., Park, Z.A. and McNabb, W.C. (2011). A comparison of analog and Next-Generation transcriptomic tools for mammalian studies. *Brief Funct Genomics,* **10**: 3, 135-150.

Schlatt, S. and Ehmcke, J. (2014). Regulation of spermatogenesis: an evolutionary biologist's perspective. *Semin Cell Dev Biol,* **29**: 2-16.

Tennessen, J.A., Bigham, A.W., O'Connor, T.D., Fu, W., Kenny, E.E*., et al.* (2012). Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science,* **337**: 6090, 64-69.

Uhlen, M., Fagerberg, L., Hallstrom, B.M., Lindskog, C., Oksvold, P*., et al.* (2015). Proteomics. Tissue-based map of the human proteome. *Science,* **347**: 6220, 1260419.

Uhlen, M., Oksvold, P., Fagerberg, L., Lundberg, E., Jonasson, K., Forsberg, M., Zwahlen, M., Kampf, C., Wester, K., Hober, S., Wernerus, H., Bjorling, L. and Ponten, F. (2010). Towards a knowledge-based Human Protein Atlas. *Nat Biotechnol,* **28**: 12, 1248-1250.

Werner, T. (2010). Next generation sequencing in functional genomics. *Brief Bioinform,* **11**: 5, 499-511.

White-Cooper, H. and Bausek, N. (2010). Evolution and spermatogenesis. *Philos Trans R Soc Lond B Biol Sci,* **365**: 1546, 1465-1480.

Wistuba, J., Stukenborg, J.-B. and Luetjens, M. (2007). Mammalian Spermatogenesis. *Functional Development and Embryology,* **1**: 99-117.

Zhu, Z., Li, C., Yang, S., Tian, R., Wang, J., Yuan, Q., Dong, H., He, Z., Wang, S. and Li, Z. (2016). Dynamics of the Transcriptome during Human Spermatogenesis: Predicting the Potential Key Genes Regulating Male Gametes Generation. *Sci Rep,* **6**: 19069.

Paper I - Autosomal Mutations and Spermatogenic Failure

Review, eLS 2014

Autosomal Mutations and Spermatogenic Failure

Ana C Lima, *Institute of Pathology and Molecular Immunology of the University of Porto (IPATIMUP), Porto, Portugal*

Alexandra M Lopes, *Institute of Pathology and Molecular Immunology of the University of Porto (IPATIMUP), Porto, Portugal*

Advanced article

Article Contents

Online posting date: 16th June 2014

Male infertility is commonly due to an impairment in the production of viable sperm, capable of fertilisation. Spermatogenic failure can manifest in its most severe form by an absence of mature sperm or, more often, by a reduction in sperm counts, as an isolated phenotype or in combination with abnormalities in sperm motility and morphology. In only a fraction of patients with primary spermatogenic failure can an underlying genetic cause be identified, such as karyotype abnormalities and Y chromosome microdeletions. However, a small number of autosomal genes harbour rare variants/mutations with strong evidence of their impact in primary spermatogenic failure in otherwise healthy men (*SYCP3*, *NR5A1*, *DMRT1*, *CATSPER*, *DNAI1*, *DNAH5*, *DNAH11*, *SLC26A8*, *AURKC*, *SPATA16*, *DPY19L2*, *KLHL10* and *SEPT12*). Even though the clinical relevance of the majority of these variants is still uncertain, they represent promising markers for spermatogenesis deficits.

Introduction

In humans, the production of male gametes initiates at puberty and is the outcome of an intricate and highly orchestrated developmental process – spermatogenesis. This process consists in the transformation of male germ cells (GCs) into mature gametes and requires a strict coordination between numerous cell types during its progression (**Figure 1**).

Given its complexity, it is predictable that abnormalities at different steps of spermatogenesis will result in a spectrum of different phenotypes affecting fertility, as supported by the panoply of mouse models of male infertility. Spermatogenic failure typically results in a decreased number of spermatozoa (SPZ; oligozoospermia), or total absence of sperm production (non-obstructive azoospermia – NOA), the latter affecting 10–15% of all patients. Some infertile patients may present abnormal forms of sperm (teratozoospermia) or decreased sperm motility (asthenozoospermia – AZS) as isolated phenotypes or in combination with low sperm counts (**Table 1**). The most frequent identifiable causes of azoospermia and severe oligozoospermia are chromosomal abnormalities, mainly Kleinefelter Syndrome (XXY) and Y chromosome microdeletions in the Azoospermia Factor regions (AZFa, AZFb and AZFc). Collectively these account for a considerable fraction of the most severe cases (30%); however, in approximately half of the patients the etiopathogenesis of spermatogenic failure remains undefined and is referred to as idiopathic infertility. **See also**: Molecular Genetics of Spermatogenic Disturbances

A few studies have analysed cohorts of patients with well-defined clinical phenotypes, and a number of convincing associations of both single nucleotide and structural variants (deletions mostly) on the autosomes and on the X chromosome have been established. Overall, some of the most promising associations obtained involve rare variants, often classified as mutations. This is unsurprising given that any highly penetrant dominant mutation with a severe impact on fertility should be rare. Recessive variants may be maintained for a longer period at a low frequency in the population and their phenotypic impact may only be revealed in ethnic groups with high levels of intra-familial marriages. **See also**: X Chromosome and Spermatogenesis Defects

The validation of potentially deleterious variants is particularly challenging in human male infertility research given that familial data is seldom available. Another limiting factor in this field, for studies focused in human, is the access to the relevant tissue which is only obtained by testicular biopsy. Also, functional assays are often hampered by the lack of a reliable *in vitro* model of spermatogenesis, therefore, highlighting the importance of mouse models. Integration of detailed clinical phenotypes and genetic data from a large number of patients, applying population genetics principles and sound statistical approaches has a crucial role in disentangling the impact of large numbers of newly discovered genetic variants on the success of spermatogenesis.

eLS subject area: Genetics & Disease

How to cite:
Lima, Ana C; and Lopes, Alexandra M (June 2014) Autosomal Mutations and Spermatogenic Failure. In: eLS. John Wiley & Sons, Ltd: Chichester.
DOI: 10.1002/9780470015902.a0025310

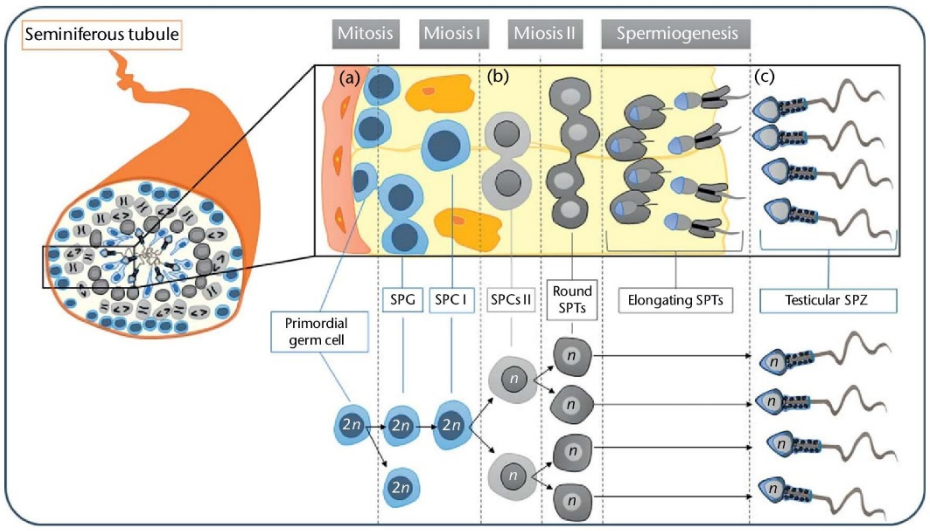


Figure 1 Overview of human spermatogenesis. In this schematics the process of spermatogenesis is depicted in an overview of a cross-section of a seminiferous tubule. Male gametes develop within the epithelia (b – Sertoli cells) of the seminiferous tubules in the testis, with differentiation occurring from the wall (a) towards the lumen (c). Diploid GCs undergo several rounds of mitosis giving rise to type A spermatogonia (SPG), to maintain the pool of stem cells and type B SPG that differentiate into primary spermatocytes (SPC I). These cells then go through meiosis I, where recombination of genetic material occurs, and originate haploid secondary spermatocytes (SPCs II). A second round of meiosis originates haploid round spermatids (SPTs) that, through a process called spermiogenesis, acquire sperm cell specialisations and are released (spermiation) into the lumen, becoming testicular SPZ. Testicular sperm cells are immotile and require the unique environment of the epididymis to gain such function. The last step of sperm maturation (capacitation) occurs within the female reproductive tract where SPZ become competent for fertilisation.

Table 1 Reference values for semen quality parameters in different pathologies[a]

Nomenclature	Sperm profile
Azoospermia	No spermatozoa (SPZ) in ejaculate
Oligozoospermia	Sperm concentration below $15 \times 10^6\,mL^{-1}$; total sperm number below 39×10^6 per ejaculate
Asthenozoospermia	Less than 32% progressively motile SPZ
Teratozoospermia	Less than 4% morphologically normal SPZ
Cryptozoospermia	SPZ absent from fresh preparations but observed in a centrifuged pellet

[a]Reproduced from World Health Organization (2010). © World Health Organization.

Autosomal Mutations and Spermatogenesis Deficits

In this review, the authors present the autosomal genes harbouring rare variants/mutations with robust evidence of their impact in primary spermatogenic failure *sensu lato* (low to nil sperm counts, morphological sperm defects and motility defects) in otherwise healthy men (non-syndromic infertility). The authors selected studies that report rare potentially deleterious variants in human genes with compatible mouse phenotype or known protein function, and meet one of the following combined criteria: (1) including cohorts of different genetic backgrounds; (2) large number of patients (evaluated taking into account the prevalence of the phenotype) and including functional assays; (3) smaller number of patients with a well-defined clinical phenotype and family data/functional assays/reporting nonsense or protein truncating mutations. Even though the clinical relevance of the vast majority of these variants is still uncertain, they represent the most promising markers for spermatogenesis deficits. A summary of the genes and variants discussed in the text is presented in **Table 2** and **Table 3**.

Alterations in Sperm Count

Oligozoospermia and azoospermia

Synaptonemal complex protein 3 (*SYCP3*)

Meiosis is a hallmark of spermatogenesis, and involves the pairing and segregation of homologous chromosomes in a diploid parental cell into haploid gametes. The

Table 2 Genes discussed in this review harbouring variants associated with abnormal sperm parameters (sperm count, motility and morphology defects)

Gene symbol	Chromosome location	Gene function	Reported variants	Model of segregation	Associated phenotypes	Evidence
SYCP3	12q23.2	Protein is part of the synaptonemal complex, essential for synapsis during meiosis	1 Nonsense	Dominant (Haploinsufficiency)	Azoospermia	Mouse phenotype; *In vitro* functional assay
NR5A1	9q33	Nuclear receptor that regulates several genes involved in the hypothalamic–pituitary steroidogenic axis and in mammalian testis development	8 Missense	Dominant (Haploinsufficiency)	Azoospermia and severe oligozoospermia	Mouse phenotype; *In vitro* functional assay
DMRT1	9p24.3	Involved in the network of sex determination across several species and major player regulating testicular differentiation in all vertebrates	Large deletions (whole gene)	Dominant (Haploinsufficiency)	Azoospermia	Mouse phenotype; present in two cohorts of patients (European ancestry and Asian)
CATSPER1	11q12.1	Member of the CATSPER voltage-gated calcium channel family that are specific to SPZ and localise to the flagellum	2 Frameshift insertions	Recessive	Asthenozoospermia (AZS)	Mouse phenotype; familial segregation
DNAI1	9p13.3	Encodes a dynein arm subunit of the intermediate chain. Structural component of the axoneme responsible for sperm tail motility	1 Missense	Dominant	AZS	Known protein function; involved in human syndromes; familial segregation
DNAH5	5p15.2	Encodes a dynein arm subunit of the heavy chain. Structural component of the axoneme responsible for sperm tail motility	1 Missense	Dominant	AZS	Known protein function; involved in human syndromes
DNAH11	7p21	Encodes a dynein arm subunit of the heavy	1 Missense	Dominant	AZS	Known protein function; involved in *(continued)*

Table 2 Continued

Gene symbol	Chromosome location	Gene function	Reported variants	Model of segregation	Associated phenotypes	Evidence
		chain. Structural component of the axoneme responsible for sperm tail motility				human syndromes; Familial segregation
SLC26A8	6p21.31	Encodes a germ cell-specific member of the SLC26 gene family of anion transporters with apparent involvement in the regulation of cystic fibrosis transmembrane conductance regulator channels. Required for proper sperm motility and capacitation	3 Missense	Dominant	AZS	Mouse phenotype; *In vitro* functional assay
AURKC	19q13	Chromosome segregation and meiotic cell division as an integral part of the spindle assembly checkpoint	1 Deletion; 1 nonsense; 1 missense; 1 splice variant	Recessive	Macrozoospermia	Known protein function; Present in several cohorts of patients (North African, European and Asian)
SPATA16	3q26.31	Putative role in acrosome formation during spermiogenesis	1 Missense	Recessive	Globozoospermia	Putative function; familial segregation
DPY19L2	12q14.2	Transmembrane protein with a role in sperm head elongation and acrosome formation	1 Recurrent whole gene deletion; 4 missense; 3 nonsense; 2 exonic deletions; 4 frameshift deletions; one splice-site mutation	Recessive	Globozoospermia	Mouse phenotype; replication in several cohorts (North African, European and Asian patients)
KLHL10	17q21.2	Protein involved in testis-specific protein ubiquitination during spermiogenesis	1 Splicing; 2 missense	Dominant (Haploinsufficiency)	Mild and severe oligozoospermia, Teratozoospermia and AZS	Mouse phenotype; *In vitro* functional assay

SEPT12	16p13.3	Encodes a testis-specific guanosine triphosphatases localised in the sperm neck and annulus	2 Missense	Dominant (haploinsufficiency/ dominant negative?)	Asthenoteratozoospermia	Mouse phenotype; *In vitro* functional assay

Table 3 Rare variants/mutations shown to be associated with abnormal sperm parameters

Gene	Variant	Patient genotype	Frequency (patients affected)	Patient phenotype	Patient ethnicity	Present in controls	Notes	Reference
SYCP3	c.643delA	Heterozygous	2/19	Azoospermia	Hispanic, Arab	No	Truncated protein at the *C*-terminal	Miyamoto *et al.* (2003)
NR5A1	p.Gly123Ala (c.368G>C)/ p.Pro129Leu (c.386C>T)	Heterozygous	3/315	Azoospermia, Oligozoospermia	Congolese, Tunisian	No	Also found in a girl with primary ovarian insufficiency[a]	Bashamboo *et al.* (2010)
	p.Pro131Leu	Heterozygous	1/315	Azoospermia	Sri Lankan	No	Within protein ligand-binding domain; impairs transactivation of NR5A1 target genes	
	p.Gly165Arg	Heterozygous	1/488	Azoospermia	Pakistanese	No	Highly conserved site in the hinge region	Ropke *et al.* (2012)
	p.Asp257Asn	Heterozygous	1/488	Cryptozoospermia	German	No	Highly conserved site in the ligand-binding domain	
	p.Ile323Thr	Heterozygous	1/488	Severe oligozoospermia	German	No	Highly conserved site in the ligand-binding domain	
	p.Arg191Cys	Heterozygous	1/315	Severe oligozoospermia	Congolese	No		
	p.Gly212Ser	Heterozygous	1/315	Severe oligozoospermia	French–Vietnamese	No		
	p.Asp238Asn	Heterozygous	1/315	Severe oligozoospermia	Egyptian	No		
DMRT1	Deletion	Heterozygous	2/726	Azoospermia	Caucasian (North American)	No	Partial or whole gene deletion	Lopes *et al.* (2013)
	Deletion	Heterozygous	3/979	Azoospermia	Han Chinese	No		
CATSPER1	Insertion	Heterozygous	2 familial cases	Asthenozoospermia (AZS)	Iranian	No	Truncated protein at the *C*-terminus	Avenarius *et al.* (2009)
DNAI1	p.Arg663Cys	Heterozygous	3/90	AZS	Italian	No	Within highly conserved sites of the protein	Zuccarello *et al.* (2008)
DNAH5	p.Glu2666Asp	Heterozygous	1/90	AZS	Italian	No		
DNAH11	p.Ile3040Val	Heterozygous	3/90	AZS	Italian	No		

SLC26A8	p.Arg87Gln	Heterozygous	1/146	AZS and incomplete annullus	Recruited in France (unspecified)	No	Mutant proteins do not activate cystic fibrosis transmembrane conductance regulator-dependent anion transport	Dirami *et al.* (2013)
	p.Glu812Lys	Heterozygous	1/146			No		
	p.Arg954Cys	Heterozygous	1/146			No		
AURKC	c.144delC	Homozygous	68/95	Macrozoospermia (with multiflagelar polyploid SPZ)	French, North African	Yes, heterozygosity	Truncated protein at the *C*-terminus (no kinase domain)	Dieterich *et al.* (2007), Dieterich *et al.* (2009), Ben Khelifa *et al.* (2012)
	Cys229Tyr	Compound heterozygous (with c.144delC)	1/41		North African	No	Conserved residue of the protein kinase domain	Dieterich *et al.* (2009)
	c.436–2A > G		1 familial case		Tunisian	No	Splice site mutation confirmed in patient messenger ribonucleic acid (mRNA)	Ben Khelifa *et al.* (2011)
	pTyr248*	Homozygous	9/44		North African, European	No	Truncated protein at the *C*-terminus	Ben Khelifa *et al.* (2012)
	pTyr248*	Compound heterozygous (with c.144delC)	2/44					
SPATA16	p.Arg283Gln	Homozygous	1 familial case	Globozoospermia	Middle Eastern	No	Within the highly conserved TPR domain	Dam *et al.* (2007)
DPY19L2	Gene deletion	Homozygous	66/144	Globozoospermia	North African, Middle Eastern, European and Chinese	Yes, heterozygosity	Recurrent deletion with several breakpoints defined	Harbuz *et al.* (2011), Koscinski *et al.* (2011), Ellnati *et al.* (2012), Coutton *et al.* (2012), Zhu *et al.* (2013)
	p.Arg290His	Compound heterozygous (with gene deletion)	3/103		European, Chinese	No, dbSNP (rare)[b]	Highly conserved residue	Ellnati *et al.* (2012), Coutton *et al.* (2012), Zhu *et al.* (2013)

(*continued*)

Table 3 Continued

Gene	Variant	Patient genotype	Frequency (patients affected)	Patient phenotype	Patient ethnicity	Present in controls	Notes	Reference
	p.Thr493Arg		1/54		Turkish	No	Predicted deleterious (PolyPhen)	Ellnati *et al.* (2012)
	p.Lys680*		1/54		Belgian	No	Premature stop codon	
	Ex5_6del		1/54		North American	No	Abnormal truncated protein	
	p.Gln345*		1/54		French	No	Premature stop codon	
	c.1183delT		2/54		Belgian	No	Frameshift, truncated protein	
	p.Gln342*		1/34		European	No	Premature stop codon	Ellnati *et al.* (2012), Coutton *et al.* (2012)
	Ex5_6del	Homozygous	1/54		French	No	Abnormal truncated protein	Ellnati *et al.* (2012)
	c.1218+1G>A		1/54		Italian	No	Splice site mutation, exon 11 skiping demonstrated *in vitro*	
	p.Arg298Cys		1/54		Moroccan	No	Highly conserved residue	
	Ex5_7del		1/54		Iranian	No	Abnormal protein lacking three transmembrane domains	
	p.Met358Lys		1/34		North African	No	Conserved residue	Coutton *et al.* (2012)
	c.1532delA		1/15		Chinese	No	Frameshift, truncated protein	Zhu *et al.* (2013)
	c.[1679delT; 1681_1682delAC] c.Leu330Pro		1/15			No	Premature stop codon	
			1/15			No	Highly conserved residue	
KLHL10	IVS3_121delTCTT	Heterozygous	1/556	Oligozoospermia	Diverse ethnicity	Not tested	mRNA with splicing defect predicted to result in an early truncated protein	Yatsenko *et al.* (2006)

	p.Gln216Pro		4/556			Yes (1/394)	Substitution in the BACK domain		
	p.Ala313Thr		2/556	Severe oligozoospermia		No	Substitution affecting the first kelch repeat motif		
SEPT12	p.Thr89Met	Heterozygous	1/160	Astenoteratozoospermic	Han Taiwanese	No	Highly conserved residue of guanosine triphosphatases (GTPases) domain	Kuo *et al.* (2012)	
	p.Asp197Asn		1/160			No	Highly conserved residue of GTPase domain		

[a]Data from Lourenço *et al.* (2009). © Massachusetts Medical Society.
[b]The Single Nucleotide polymorphism Database (dbSNP).

synaptonemal complex, composed of several proteins, is essential for synapsis of homologous chromosomes and successful meiosis (Heyting, 1996). Male mice heterozygous for a null mutation in *Sycp3*, a structural component of the synaptonemal complex, are sterile due to spermatogenesis arrest at meiotic prophase (Yuan *et al.*, 2000). *SYCP3* (OMIM604759) is, therefore, a convincing candidate for association with azoospermia caused by maturation arrest at meiosis.

In the first report (Miyamoto *et al.*, 2003) the same 1 bp deletion was found in *SYCP3* coding region, in two men of different ethnicity (one Hispanic and one Arab), out of a group of 19 North American azoospermic patients with maturation arrest. This deletion (c.643delA) results in a premature stop codon and a truncated SYCP3 protein at the *C*-terminal, which displayed reduced interaction with the wild-type protein *in vitro* and interfered with SYCP3 fibre formation in cultured cells. None of the 75 fertile controls analysed carried the deletion. In spite of these encouraging initial results, three following studies in 85 central European, 13 Arabic patients (Stouffs *et al.*, 2005, 2011) and 22 Mediterranean patients (Martínez *et al.*, 2007) did not find patient-specific deleterious mutations in *SYCP3* mutations in low sperm counts or in azoospermia with maturation arrest.

The difference in prevalence of *SYCP3* mutations between studies is most likely related to the genetic background of the populations analysed and indicate that larger studies comprising different populations are needed to fully assess the impact of these variants.

Nuclear receptor subfamily 5, group A, member 1 (*NR5A1*)

NR5A1 (OMIM184757), also called Steroidogenic Factor 1 (SF1) is an orphan nuclear receptor with a crucial role in the development of the adrenal glands and the gonads. In mice, SF1 is expressed in the embryonic urogenital ridge at early stages of gonadal development, in foetal Sertoli cells and in primary steroidogenic tissues of the male adult – adrenal cortex and Leydig cells (Ikeda *et al.*, 1994). In knockouts of the *NR5A1* orthologue (*Ftzf1*) adrenal and gonadal development is compromised, causing XY sex reversal (Luo *et al.*, 1994). SF1 is a key transcriptional regulator of several genes involved in the hypothalamic–pituitary steroidogenic axis and in mammalian testis determination and differentiation, as a positive regulator of SOX9 (Sry-box 9) and anti-Müllerian hormone (AMH) (Lin and Achermann, 2008; Sekido and Lovell-Badge, 2008). The protein encoded by *NR5A1* consists of a deoxyribonucleic acid (DNA)-binding domain including two zinc fingers, a flexible hinge region, a ligand-binding domain and two activation domains.

NR5A1 mutations were initially associated with more severe disorders of sex development (DSD) including 46, XY partial and complete gonadal dysgenesis with or without adrenal failure (Lin *et al.*, 2007). More recently, 315 men with idiopathic spermatogenic failure recruited in France were analysed (Bashamboo *et al.*, 2010) and six heterozygous missense mutations were found within *NR5A1*, in patients with moderate to severe reduction in sperm counts, all of whom had mixed non-European ancestry. Three men of African origin carried a double *NR5A1* mutation (p.Gly123Ala/p.Pro129Leu; NP_004950) that is probably a founder mutation. All these mutations were located within functional domains of the protein involved in ligand-binding stabilisation and four of them (Pro131Lys, Arg191Cys, Gly212Ser and Asp238Asn) were shown to alter the capacity of the protein to transactivate two of its targets (AMH and Cyp11a1). None of the mutations was detected in more than 2100 control individuals, including 370 fertile men and 359 normozoospermic men. Additionally, three heterozygous mutations predicted to be damaging to the protein were identified in *NR5A1* in a large study including 488 German patients with azoo- or severe oligozoospermia (Ropke *et al.*, 2013). Therefore, *NR5A1* mutations seem to contribute to a fraction of cases of severe spermatogenic failure (SSF) in different ethnic backgrounds, even though the prevalence of mutations may not be as high in Europeans as initially found in North Africans.

Doublesex and mab-3 related transcription factor 1 (*DMRT1*)

DMRT genes contain a DM (Doublesex and mab-3) domain DNA-binding motif known to regulate sex determination across several species. Particularly, the *DMRT1* gene (OMIM602424) is a major regulator of testicular differentiation in all vertebrates (Zarkower, 2013). Conditional knockout of this gene in GCs of adult mice testis (Matson *et al.*, 2011) have revealed impaired GCs development. In humans, its expression pattern is testis-specific in the adults, with transcripts being detected in both Sertoli and germ cells. All the data is suggestive of DMRT1 playing a critical role in male gonadal development and maintenance. In accordance, several syndromic human phenotypes have been associated with the loss of DMRT1 supporting a model of haploinsufficiency for this gene in humans. Hemizygosity of *DMRT1* is associated with XY sex reversal (Matson and Zarkower, 2012) and linked with DSD such as gonadal dysgenesis (Tannour-Louet *et al.*, 2010). Interestingly, in a recent genome-wide analysis of copy number variants (CNVs) in patients with SSF (Lopes *et al.*, 2013) two microdeletions spanning *DMRT1* coding sequence were found in two azoospermic unrelated Caucasian men recruited in Utah, one with Sertoli cell only (SCO) syndrome and other showing spermatocytic arrest. Overall the authors identified a total of five deletions of *DMRT1* coding sequence in cases of European and Chinese ancestry and none in greater than 7000 controls. From these combined data the frequency of *DMRT1* exonic deletion was estimated: 0.38% (5/1306) in cases and 0% (0/2858) in controls (OR = Infinity, [2.0−Inf], $p = 0.003$). In the light of these and previous findings it is safe to state that the presence of these apparently globally dispersed albeit rare deletions has a

clear impact on male fertility, even though their penetrance may be variable, likely depending on the size and content of the deleted region and on the genetic background of the individual. The search for mutations within *DMRT1* coding and regulatory region in European azoospermic patients suggests that variants within this gene may also contribute to spermatogenic failure even though they are rare (Lima, personal communication).

Defects in Sperm Motility

Asthenozoospermia

AZS is characterised by decreased sperm motility (<32% motile SPZ) and is a common cause of male infertility. A few single gene defects have been convincingly associated with non-syndromic AZS, including mutations in cation channels and solute transporters (*CATSPER1* and *SLC26A8*) and axonemal dynein genes (*DNAI1*, *DNAH5* and *DNAH11*).

Cation channel, sperm associated 1 (*CATSPER1*)

To become competent to fertilise the egg, sperm have to undergo a series of biochemical events once in the female reproductive tract, in a process named capacitation, which depends on changes in intracellular Ca^{2+}. This gene is one of four members of the sperm-specific CATSPER voltage-gated calcium channel family shown to be essential for male fertility in mice (Ren *et al.*, 2001; Carlson *et al.*, 2003; Qi *et al.*, 2007). It is specifically expressed in the plasma membrane of SPZ above the fibrous sheath in the principal piece of the mature sperm tail (Ren *et al.*, 2001; **Figure 2**).

Two different frame shifting insertions in the first exon of *CATSPER1* (OMIM606389) gene were identified in homozygosity in two consanguineous Iranian families (Avenarius *et al.*, 2009). The males carrying these insertions were infertile and presented sperm defects suggestive of AZS. None of the 576 geographically matched controls carried either of the insertions. The presence of these mutations is expected to result in severely truncated proteins lacking all transmembrane domains and the channel pore, if not degraded by messenger ribonucleic acid (mRNA) nonsense-mediated decay. These results suggest that CATSPER ion channels are essential for fertility in humans and *CATSPER1* insertions constitute an example of recessive Mendelian inheritance of non-syndromic male infertility.

Solute carrier family 26 (anion exchanger), member 8 (*SLC26A8*)

SLC26A8 (OMIM608480) is part of the solute-linked carrier 26 (SLC26) family of anion transporters and is restricted to male GCs. Mice that are *Slc26A8*-null are sterile and show complete lack of sperm motility, impairment of sperm capacitation and severe structural defects of the flagellum (Touré *et al.*, 2007). Human SLC26A8 can be detected in mature sperm, in the equatorial segment of the head and in the annulus, a ring-shaped structure at the junction between the mid piece and the principal piece of the flagellum (Touré *et al.*, 2007; Rode *et al.*, 2012).

Dirami *et al.* (2013) have screened by high-performance liquid chromatography and sequencing 146 infertile men with AZS, including 27 severe cases, and found 7 *SLC26A8* coding variants in heterozygosity that were absent from 121 controls matched for ethnicity. Given that SLC26A8 forms a complex with the cystic fibrosis transmembrane

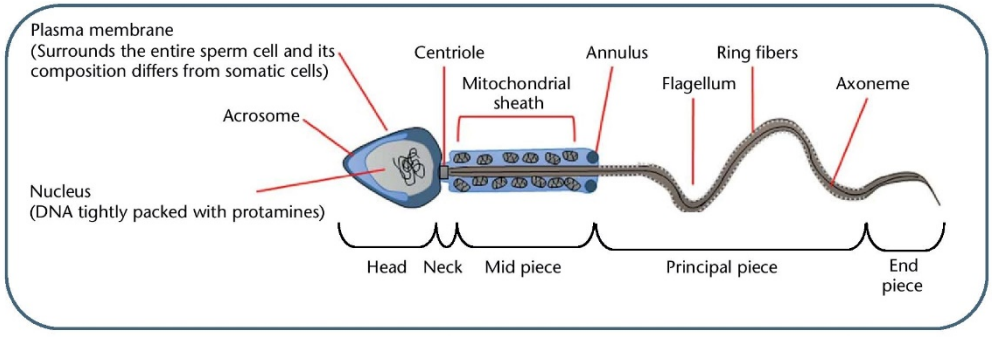


Figure 2 Diagram of longitudinal section of a human spermatozoon. Human male gametes are highly specialized cells with a structure optimised for fertilisation. Sperm cells lack many cytoplasmatic organelles, such as endoplasmic reticulum, Golgi apparatus or ribosomes, but develop a unique secretory vesicle called acrosome which contains many hydrolytic enzymes required for egg penetration. In the nucleus, histones are replaced by protamines to allow DNA compaction into a much reduced volume. The capacity to move in an aqueous medium is provided by a strong flagellum. The mid piece is the region of the cell that powers the flagellum for motility by the presence of a mitochondrial sheath displayed as a spiral structure around the axoneme, resulting from the fusion of individual mitochondria during SPT differentiation. The axoneme is a cytoskeleton structure in the inner core of the flagellum, going from the proximal centriole until the end of the tail. In mammals, it consists of two singlet microtubules surrounded by nine microtubule doublets and due to this distinctive organisation and the presence of axonemal dyneins, is the structure responsible for generating movement of the flagellum. The unique organisation and specialisation of cellular structures in the spermatozoon allows its hypermotility and the proper fertilisation of the female egg.

conductance regulator (CFTR), also required for sperm motility and capacitation (Rode *et al.*, 2012), the authors tested *in vitro* the effect of each mutation in the interaction between the two proteins. They demonstrated that only three of these sequence variants (p.Arg87Gln, p.Glu812Lys and p.Arg954Cys) were deleterious, since mutant proteins did not activate CFTR-dependent anion transport, even though the interaction between the two proteins was partially retained. Additionally they found that the sperm of patients had aberrantly low levels of SLC26A8 protein and morphological abnormalities, such as midpiece disorganisation and an incomplete annulus, similar to the defects observed in *Slc26A8*-null sperm in mice.

These findings indicate that the function of the SLC26A8–CFTR complex is compromised in these patients leading to a decreased motility of the SPZ, and suggest a role for this complex in the motility and capacitation of the male mature gamete.

Dynein genes: dynein, axonemal, intermediate chain 1 (*DNAI1*), dynein, axonemal, heavy chain 5 (*DNAH5*) and dynein, axonemal, heavy chain 11 (*DNAH11*)

Dyneins are intracellular motor proteins that convert the energy contained in adenoside triphosphate in mechanical energy, allowing them to literally move. They transport their cellular cargo by 'walking' along the microtubules within the cells. Briefly, there are two classes of dyneins: cytoplasmatic dyneins perform functions necessary for cell survival such as organelle transport and centrosome assembly; and axonemal dyneins, only found in cells containing motile structures, cause the sliding of microtubules in the axonemes of cilia and sperm tail flagella promoting their movement. The axoneme is an evolutionary conserved structure and consists of several microtubules in a precise structural concentric arrangement (9 + 2). Structural defects of axonemal dyneins are present in a large proportion of patients with syndromes that include sperm motility defects, such as primary ciliary dyskinesia (PCD, OMIM242650) and Kartagener syndrome (OMIM244400), and mutations in three of the several dynein arms subunits have been identified in these patients *DNAI1* (OMIM604366), *DNAH5* (OMIM603335) and *DNAH11* (OMIM603339). A screen in 90 Italian men with isolated non-syndromic AZS revealed mutations in 7 patients, in each of the three dynein genes known to be affected in PCD (Zuccarello *et al.*, 2008). These mutations, affecting highly conserved residues of the encoded proteins, were absent from 200 normospermic controls. In two consanguineous families the *DNAI1* and *DNAH11* mutations found had been transmitted to the infertile patients by their healthy mothers. However, no ultrastructural axonemal anomaly was detected in the sperm of these patients. The segregation pattern observed in these patients may imply an autosomal dominant mode of transmission and suggests that variants affecting protein function might be kept in the population at low frequency by female carriers.

Teratozoospermia – Abnormal Sperm Morphology

Macrozoospermia

Aurora kinase C (*AURKC*)

Male infertility associated with large-headed, multiflagellar and polyploid SPZ (OMIM243060), a phenotype mainly observed in North Africans, is caused by homozygous or compound heterozygous mutations in the *AURKC* (OMIM603495) gene. Overall four mutations have been identified: one frameshift deletion, one recurrent nonsense mutation found in several unrelated individuals, a missense mutation observed in only one patient and a splicing variant resulting in loss of an exon, identified in two brothers. Research carried out on AURKC, an aurora kinase preferentially expressed in testis, demonstrated its essential role in chromosome segregation and meiotic cell division (Tang *et al.*, 2006). This agrees well with the phenotype of meiotic arrest observed in macrozoospermia with *AURKC* inactivating mutations resulting in aneuploid gametes.

Dieterich *et al.* (2007) carried out a whole genome scan in 10 infertile patients of North African origin with 100% of sperm with typical morphological aberrations and increased chromosomal content, which led to the discovery of the same homozygous mutation on *AURKC* (c.144delC) in all the patients tested. This mutation resulted in a premature translation termination and in a non-functional protein without a kinase domain. Even though four of the patients were unrelated French citizens and the remaining six came from Morocco they all shared the same haplotype and therefore had a common ancestor. In a follow up study (Dieterich *et al.*, 2009), 62 patients with North African ancestry were genotyped and all who had a typical phenotype with close to 100% large-headed SPZ were homozygous for the c.144delC ($n = 32$), whereas no *AURKC* mutations were detected in the others. According to this study, in the Maghrebian general population this founder mutation reaches 1% frequency, increasing, therefore, the risk of teratozoospermia deriving either by c.144delC homozygosity or from compound heterozygosity. Concordantly, one new mutation in a conserved residue of the protein kinase domain (Cys229Tyr), was identified in a compound heterozygote for the deletion. In all patients homozygous for *AURKC* mutations, higher DNA content was detected thus proving that the development of their gametes arrested before the first meiotic division.

The genotype–phenotype correlation for *AURKC* genetic alterations was further confirmed in two later studies with North African and European patients. Following the characterisation of two brothers of Tunisian descent with high percentage of macrozoospermia who were compound heterozygotes for a new heterozygous splicing mutation and the c.144delC (Ben Khelifa *et al.*, 2011), the same authors found a second nonsense mutation (p.Y248*) in unrelated individuals of European and North African

origin (Ben Khelifa *et al.*, 2012). This alteration represented 13% of all AURKC mutant alleles detected in a total of 87 infertile men of mostly North African ancestry analysed in this study, all presenting more than 5% large-headed SPZ. The phenotype associated with the (p.Y248*) mutation was indistinguishable from that initially described in individuals carrying the main AURKC mutation (c.144delC). Concordantly, sperm parameters of patients with no identified *AURKC* mutation were significantly better, with a decreased rate of SPZ with a large head and multiple flagella. A recent study confirmed the high prevalence of the c.144delC in Morrocan infertile patients and the genotype–phenotype association for homozygotes (Eloualid *et al.*, 2014). Overall these results confirm that AUKRC deficiency is among the most frequent single cause of male infertility in Maghrebian men.

Globozoospermia

Globozoospermia (OMIM613958) is a rare phenotype of primary male infertility (<0.1%) characterised by malformation or absence of the acrosome and round-headed SPZ. Two autosomal genes have been robustly associated with this phenotype, through the finding of homozygous mutations in families and in several unrelated patients: *SPATA16* (OMIM609856; 3q26.31) and *DPY19L2* (OMIM613893).

Spermatogenesis-associated 16

Dam *et al.* (2007) have studied an Ashkenazi Jewish family with three affected brothers and, by analysing a shared region of homozygosity, detected a missense mutation in a highly conserved residue of spermatogenesis-associated 16 (SPATA16), also known as 'NYD-SP12'. This was one of the first reports of an autosomal mutation in a single gene segregating with a non-syndromic male infertility condition. The affected testis-specific gene (Xu *et al.*, 2003) encodes a protein that migrates to the acrosome with Golgi vesicles in round and elongated SPTs, strongly suggesting a role for the SPATA16 protein in acrosome formation during spermiogenesis (Lu *et al.*, 2006). No other mutations in *SPATA16* have been described in 29 more patients included in the initial study (Dam *et al.*, 2007).

Dpy-19-like 2 (*DPY19L2*)

DPY19L2 is a member of the *DPY19L* family, which arose by gene duplication and whose function in spermatogenesis was unsuspected. A homozygous deletion spanning *DPY19L2* was initially described in 75% of patients with North African ancestry analysed, mainly from Tunisia (Harbuz *et al.*, 2011) and in four out of 24 patients, including a consanguineous family and 3 unrelated men recruited from Europe and the Middle East (Koscinski *et al.*, 2011). A subsequent study including a total of 54 unrelated patients from 13 different countries within Europe and North Africa confirmed the high prevalence of *DPY19L2* mutations in patients with globozoospermia – 66.7% (ElInati *et al.*, 2012). The *DPY19L2* recurrent deletion is driven by non-allelic homologous recombination between two highly homologous low copy repeats at 12q14.2 and overlaps a CNV with duplications and heterozygous deletions in control individuals from different ethnic groups. However, none of the, approximately 500, fertile controls of both North African and European origin screened in the aforementioned studies were homozygous for the deletion, indicating that only one copy of functional protein is sufficient to avoid the development of sperm abnormalities. Altogether, studies involving 165 patients, including several familial cases from different ethnic backgrounds and geographic regions (Europe, Middle East, North Africa and Asia) confirmed that *DPY19L2* deletion is the mutation responsible for the majority of globozoospermia cases, either in homozygosity or as a second lesion in compound heterozygotes with *DPY19L2* point mutations (Harbuz *et al.*, 2011; Koscinski *et al.*, 2011; Coutton *et al.*, 2012; ElInati *et al.*, 2012; Zhu *et al.*, 2013). Functional studies in mice have established a role for this transmembrane protein in sperm head elongation and acrosome formation (Pierre *et al.*, 2012). Moreover, the existence of additional genetic alterations that disrupt the *DPY19L2* coding sequence, including several different missense substitutions, frameshifts and nonsense mutations (**Table 3**), either in homozygous state or in compound heterozygosity with the deletion, reinforce its importance in human spermiogenesis. These findings will certainly allow the development of a diagnostic test for a large fraction of the patients with globozoospermia.

Mixed Phenotypes

Sperm count, motility and morphological defects

Kelch-like family member 10 (*KLHL10*)

The disruption of mouse ortholog of *KLHL10* (OMIM608778), a testis-specific protein that contains a BTB (BR-C, ttk and bab-bric à brac)/POZ (Pox virus and Zinc finger) domain and six kelch repeats, represents one of the few known models of autosomal dominant male infertility (Yan *et al.*, 2004). This protein has been shown to interact with CUL3 to form a CUL3-based ubiquitin E3 ligase (Wang *et al.*, 2006) that functions specifically in the testis to mediate protein ubiquitination during spermiogenesis.

Using an unconventional approach, Yatsenko *et al.* (2006) discovered three heterozygous alterations in *KLHL10* in 7 out of 556 oligozoospermic patients (1.3%) of diverse ethnicity recruited at a clinic in Houston, Texas. Interestingly, by sequencing both sperm mRNA and genomic DNA the authors were able to detect in one of the patients with severe oligozoospermia a splicing defect that is predicted to result in an early truncated protein, and a likely causal intronic deletion (IVS3_121delTCTT). Additionally, two missense alterations involving highly

conserved residues of KLHL10 were discovered (**Table 3**): the Gln216Pro in the BACK (BTB and C-terminal Kelch) domain of the protein and a second substitution affecting the first kelch repeat motif of the protein (Ala313Thr). Besides oligozoospermia, 5 of 7 patients with *KLHL10* mutations showed teratozoospermia and 4 of 7 patients had moderate AZS. None of the 394 geographically matched normozoospermic controls had the Ala313Thr substitution and only one harboured the Gln216Pro substitution. Notably, in yeast interaction assays both mutant proteins (Gln216Pro and Ala313Thr) display impaired homodimerization with the wild-type protein suggesting a functional impact. These findings are in agreement with the phenotype of mice heterozygous for a *Klhl10* defect, which are infertile and display abnormal SPT maturation and severe oligozoospermia (Yan *et al.*, 2004).

When patients and controls analysed (Yatsenko *et al.*, 2006) were matched for Y haplogroup, to correct for population structure, only the alterations found in severe oligozoospermic patients (defined as having $< 10 \times 10^6$ sperm mL^{-1}) showed a significant association. In spite of disrupting the protein function, given the high cut-off used for oligozoospermia in this study ($< 34 \times 10^6$ sperm mL^{-1}) and the asymmetric distribution of the Gln216Pro alteration in different populations ($\sim 1\%$ frequency in African Americans and absent in European Americans according to recent data from the Exome Sequencing Project (ESP)), a larger and more restrictive survey in patients of European ancestry is necessary to establish its impact in male infertility. Neither of these alterations have been found in a large cohort of, approximately 300, Chinese patients (Qiu *et al.*, 2009) with low sperm counts (azoo- and oligozoospermic) or sperm motility defects (asthenozoospermic) and no other *KLHL10* missense alterations have been reported. Overall, evidence from mice and the presence of a splicing defect and damaging missense heterozygous substitutions in human *KLHL10* support a model of haploinsuficiency for this gene in male infertility, even though replication of in other populations is still essential.

Septin 12 (*SEPT12*)

SEPT12 is a testis-specific septin and belongs to a conserved family of cytoskeletal guanosine triphosphatases (GTPases), proteins that through GTP binding and hydrolysis regulate diverse physiological processes. Lin *et al.* (2006) first found that *SEPT12* (OMIM611562) was downregulated in men with azoospermia by maturation arrest and SCO, and later showed by immunofluorescence that the encoded protein was present in sperm neck and annulus (Lin *et al.*, 2011). Moreover, Sept12 +/– chimeric mice are sterile and show various sperm defects, including immotility, bent tail, acrosome break and round head (Lin *et al.*, 2009, 2011).

Following a sequencing survey in 160 Han Taiwanese infertile patients, Kuo *et al.* (2012) found two heterozygous missense mutations in highly conserved residues of SEPT12 GTPase domain (Thr89Met and Asp197Asn), in two patients showing abnormal sperm motility and morphology (asthenoteratozoospermic). None of the mutations was present in any of the 200 fertile controls analysed, all with normal sperm parameters. The authors performed several functional assays and showed that both mutations affected the GTPase function of the protein and interfered with wild-type SEPT12, in a dosage-dependent manner. Remarkably, the sperm of the patient harbouring the Asp197Asn mutation had multiple structural defects, including a defective annulus with bent tail and no SEPT12 signal was detected in this structure. Recently, the same group showed that SEPT12 is necessary for microtubule organisation during mouse spermiogenesis (Kuo *et al.*, 2013). The interesting finding of association of SEPT12 coding variants with defects in sperm motility and morphology agrees well with the role of this protein in spermiogenesis and awaits replication in other cohorts.

Final Remarks

The ethiopathogenesis of human spermatogenic failure is far from being defined. There is still a long road ahead to unveil the genes involved and the molecular mechanisms underlying this phenotype. Nevertheless, from the reports in the literature it becomes evident that this phenotype can arise from mutations of several different loci. Therefore, SSF may represent a common phenotype arising from a collection of rare variants with asymmetric distribution between populations that act in a Mendelian fashion, even though in many cases penetrance and expressivity may be incomplete. Future studies focusing on families with severe and well-defined phenotypes will potentially reveal new highly penetrant alterations. Additionally genome-wide approaches will certainly be valuable in pointing new candidate genes in infertility phenotypes with high genetic heterogeneity, as seems to be the case for sperm count abnormalities. New technological developments and the reduction of sequencing costs will provide opportunities for researchers to overcome some of the challenges of the male infertility field and most likely play a major role in dissecting the genetic factors underlying human spermatogenic failure.

References

Avenarius M, Hildebrand MS, Zhang Y *et al.* (2009) Human male infertility caused by mutations in the CATSPER1 channel protein. *American Journal of Human Genetics* **84**(4): 505–510.

Bashamboo A, Ferraz-de-Souza B, Lourenço D *et al.* (2010) Human male infertility associated with mutations in NR5A1 encoding steroidogenic factor 1. *American Journal of Human Genetics* **87**(4): 505–512.

Ben Khelifa M, Coutton C, Blum MGB *et al.* (2012) Identification of a new recurrent Aurora Kinase C mutation in both European and African men with macrozoospermia. *Human Reproduction* **27**(11): 3337–3346.

Ben Khelifa M, Zouari R, Harbuz R *et al.* (2011) A new AURKC mutation causing macrozoospermia: implications for human spermatogenesis and clinical diagnosis. *Molecular Human Reproduction* **17**(12): 762–768.

Carlson AE, Westenbroek RE, Quill T *et al.* (2003) CatSper1 required for evoked Ca^{2+} entry and control of flagellar function in sperm. *Proceedings of the National Academy of Sciences of the USA* **100**(25): 14864–14868.

Coutton C, Zouari R, Abada F *et al.* (2012) MLPA and sequence analysis of DPY19L2 reveals point mutations causing globozoospermia. *Human Reproduction* **27**(8): 2549–2558.

Dam AHDM, Koscinski I, Kremer JAM *et al.* (2007) Homozygous mutation in SPATA16 is associated with male infertility in human globozoospermia. *American Journal of Human Genetics* **81**(4): 813–820.

Dieterich K, Soto Rifo R, Faure AK *et al.* (2007) Homozygous mutation of AURKC yields large-headed polyploid spermatozoa and causes male infertility. *Nature Genetics* **39**(5): 661–665.

Dieterich K, Zouari R, Harbuz R *et al.* (2009) The Aurora Kinase C c.144delC mutation causes meiosis I arrest in men and is frequent in the North African population. *Human Molecular Genetics* **18**(7): 1301–1309.

Dirami T, Rode B, Jollivet M *et al.* (2013) Missense mutations in SLC26A8, Encoding a sperm-specific activator of CFTR, are associated with human asthenozoospermia. *American Journal of Human Genetics* **92**(5): 760–766.

ElInati E, Kuentz P, Redin C *et al.* (2012) Globozoospermia is mainly due to DPY19L2 deletion via non-allelic homologous recombination involving two recombination hotspots. *Human Molecular Genetics* **21**(16): 3695–3702.

Eloualid A, Rouba H, Rhaissi H *et al.* (2014) Prevalence of the Aurora kinase C c.144delC mutation in infertile Moroccan men. *Fertility and Sterility* **101**(4): 1086–1090.

Harbuz R, Zouari R, Pierre V *et al.* (2011) A recurrent deletion of DPY19L2 causes infertility in man by blocking sperm head elongation and acrosome formation. *American Journal of Human Genetics* **88**(3): 351–361.

Heyting C (1996) Synaptonemal complexes: structure and function. *Current Opinion in Cell Biology* **8**(3): 389–396.

Ikeda Y, Shen WH, Ingraham HA and Parker KL (1994) Developmental expression of mouse steroidogenic factor-1, an essential regulator of the steroid hydroxylases. *Molecular Endocrinology* **8**(5): 654–662.

Koscinski I, ElInati E, Fossard C *et al.* (2011) DPY19L2 deletion as a major cause of globozoospermia. *American Journal of Human Genetics* **88**(3): 344–350.

Kuo P-L, Chiang H-S, Wang Y-Y *et al.* (2013) SEPT12-microtubule complexes are required for sperm head and tail formation. *International Journal of Molecular Sciences* **14**(11): 22102–22116.

Kuo Y-C, Lin Y-H, Chen H-I *et al.* (2012) SEPT12 mutations cause male infertility with defective sperm annulus. *Human Mutation* **33**(4): 710–719.

Lin L and Achermann JC (2008) Steroidogenic factor-1 *(SF-1, Ad4BP, NR5A1)* and disorders of testis development. *Sexual Development* **2**(4–5): 200–209.

Lin L, Philibert P, Ferraz-de-Souza B *et al.* (2007) Heterozygous missense mutations in steroidogenic factor 1 (SF1/Ad4BP, NR5A1) are associated with 46,XY disorders of sex development with normal adrenal function. *Journal of Clinical Endocrinology & Metabolism* **92**(3): 991–999.

Lin Y-H, Chou C-K, Hung Y-C *et al.* (2011) SEPT12 deficiency causes sperm nucleus damage and developmental arrest of preimplantation embryos. *Fertility and Sterility* **95**(1): 363–365.

Lin Y-H, Lin Y-M, Teng Y-N *et al.* (2006) Identification of ten novel genes involved in human spermatogenesis by microarray analysis of testicular tissue. *Fertility and Sterility* **86**(6): 1650–1658.

Lin Y-H, Lin Y-M, Wang Y-Y *et al.* (2009) The expression level of septin12 is critical for spermiogenesis. *American Journal of Pathology* **174**(5): 1857–1868.

Lopes AM, Aston KI, Thompson E *et al.* (2013). Human spermatogenic failure purges deleterious mutation load from the autosomes and both sex chromosomes, including the gene *DMRT1*. *PLoS Genetics* **9**(3): e1003349.

Lourenço D, Brauner R, Lin L *et al.* (2009). Mutations in NR5A1 associated with ovarian insufficiency. *New England Journal of Medicine* **360**(12): 1200–1210.

Lu L, Lin M, Xu M, Zhou ZM and Sha JH (2006) Gene functional research using polyethylenimine-mediated *in vivo* gene transfection into mouse spermatogenic cells. *Asian Journal of Andrology* **8**(1): 53–59.

Luo X, Ikeda Y and Parker KL (1994) A cell-specific nuclear receptor is essential for adrenal and gonadal development and sexual differentiation. *Cell* **77**(4): 481–490.

Martínez J, Bonache S, Carvajal A, Bassas L and Larriba S (2007) Mutations of SYCP3 are rare in infertile Spanish men with meiotic arrest. *Fertility and Sterility* **88**(4): 988–989.

Matson CK, Murphy MW, Sarver AL *et al.* (2011) DMRT1 prevents female reprogramming in the postnatal mammalian testis. *Nature* **476**(7358): 101–104.

Matson CK and Zarkower D (2012) Sex and the singular DM domain: insights into sexual regulation, evolution and plasticity. *Nature Reviews Genetics* **13**(3): 163–174.

Miyamoto T, Hasuike S, Yogev L *et al.* (2003) Azoospermia in patients heterozygous for a mutation in SYCP3. *Lancet* **362**(9397): 1714–1719.

Pierre V, Martinez G, Coutton C *et al.* (2012) Absence of Dpy19l2, a new inner nuclear membrane protein, causes globozoospermia in mice by preventing the anchoring of the acrosome to the nucleus. *Development* **139**(16): 2955–2965.

Qi H, Moran MM, Navarro B *et al.* (2007) All four CatSper ion channel proteins are required for male fertility and sperm cell hyperactivated motility. *Proceedings of the National Academy of Sciences of the USA* **104**(4): 1219–1223.

Qiu QM, Liu G, Li WN *et al.* (2009) Mutation of KLHL-10 in idiopathic infertile males with azoospermia, oligospermia or asthenospermia. *Zhonghua Nan Ke Xue* **15**(11): 974–979.

Ren D, Navarro B, Perez G *et al.* (2001) A sperm ion channel required for sperm motility and male fertility. *Nature* **413**(6856): 603–609.

Rode B, Dirami T, Bakouh N *et al.* (2012) The testis anion transporter TAT1 (SLC26A8) physically and functionally interacts with the cystic fibrosis transmembrane conductance regulator channel: a potential role during sperm capacitation. *Human Molecular Genetics* **21**(6): 1287–1298.

Ropke A, Tewes AC, Gromoll J *et al.* (2013) Comprehensive sequence analysis of the NR5A1 gene encoding steroidogenic factor 1 in a large group of infertile males. *European Journal of Human Genetics* **21**(9): 1012–1015.

Sekido R and Lovell-Badge R (2008) Sex determination involves synergistic action of SRY and SF1 on a specific Sox9 enhancer. *Nature* **453**(7197): 930–934.

Stouffs K, Lissens W, Tournaye H, Van Steirteghem A and Liebaers I (2005) SYCP3 mutations are uncommon in patients with azoospermia. *Fertility and Sterility* **84**(4): 1019–1020.

Stouffs K, Vandermaelen D, Tournaye H, Liebaers I and Lissens W (2011) Mutation analysis of three genes in patients with maturation arrest of spermatogenesis and couples with recurrent miscarriages. *Reproductive Biomedicine Online* **22**(1): 65–71.

Tang C-JC, Lin C-Y and Tang TK (2006) Dynamic localization and functional implications of Aurora-C kinase during male mouse meiosis. *Developmental Biology* **290**(2): 398–410.

Tannour-Louet M, Han S, Corbett ST *et al.* (2010) Identification of *De Novo* copy number variants associated with human disorders of sexual development. *PLoS One* **5**(10): e15392.

Touré A, Lhuillier P, Gossen JA *et al.* (2007) The testis anion transporter 1 (Slc26a8) is required for sperm terminal differentiation and male fertility in the mouse. *Human Molecular Genetics* **16**(15): 1783–1793.

Wang S, Zheng H, Esaki Y, Kelly F and Yan W (2006) Cullin3 is a KLHL10-interacting protein preferentially expressed during late spermiogenesis. *Biology of Reproduction* **74**(1): 102–108.

World Health Organization (2010) *Who Laboratory Manual for the Examination and Processing of Human Semen*, 5th edn. Geneva, Switzerland: WHO Press.

Xu M, Xiao J, Chen J *et al.* (2003) Identification and characterization of a novel human testis-specific Golgi protein, NYD-SP12. *Molecular Human Reproduction* **9**(1): 9–17.

Yan W, Ma L, Burns KH and Matzuk MM (2004) Haploinsufficiency of kelch-like protein homolog 10 causes infertility in male mice. *Proceedings of the National Academy of Sciences of the USA* **101**(20): 7793–7798.

Yatsenko AN, Roy A, Chen R *et al.* (2006) Non-invasive genetic diagnosis of male infertility using spermatozoal RNA: KLHL10 mutations in oligozoospermic patients impair homodimerization. *Human Molecular Genetics* **15**(23): 3411–3419.

Yuan L, Liu J-G, Zhao J *et al.* (2000). The murine SCP3 gene is required for synaptonemal complex assembly, chromosome synapsis, and male fertility. *Molecular Cell* **5**(1): 73–83.

Zarkower D (2013) Chapter twelve – DMRT genes in vertebrate gametogenesis. In: Paul MW (ed.) *Current Topics in Developmental Biology*, vol. 102, pp. 327–356. Academic Press.

Zhu F, Gong F, Lin G and Lu G (2013) DPY19L2 gene mutations are a major cause of globozoospermia: identification of three novel point mutations. *Molecular Human Reproduction* **19**(6): 395–404.

Zuccarello D, Ferlin A, Cazzadore C *et al.* (2008) Mutations in dynein genes in patients affected by isolated non-syndromic asthenozoospermia. *Human Reproduction* **23**(8): 1957–1962.

Further Reading

Hildebrand MS, Avenarius MR, Fellous M *et al.* (2010). Genetic male infertility and mutation of CATSPER ion channels. *European Journal of Human Genetics* **18**(11): 1178–1184.

Hwang K, Yatsenko AN, Jorgez CJ *et al.* (2010). Mendelian genetics of male infertility. *Annals of the New York Academy of Sciences* **1214**(1): E1–E17.

Krausz C (2011) Male infertility: pathogenesis and clinical diagnosis. *Best Practice & Research Clinical Endocrinology & Metabolism* **25**(2): 271–285.

Massart A, Lissens W, Tournaye H and Stouffs K (2012) Genetic causes of spermatogenic failure. *Asian Journal of Andrology* **14**(1): 40–48.

Matzuk M and Lamb D (2002) Genetic dissection of mammalian fertility pathways. *Nature Cell Biology* **4**(suppl.): s41–s49.

Matzuk M and Lamb D (2008) The biology of infertility: research advances and clinical challenges. *Nature Cell Biology* **14**(11): 1197–1213.

Navarro B, Kirichok Y, Chung JJ and Clapham DE (2008) Ion channels that control fertility in mammalian spermatozoa. *International Journal of Developmental Biology* **52**(5–6): 607–613.

Nuti F and Krausz C (2008) Gene polymorphisms/mutations relevant to abnormal spermatogenesis. *Reproductive BioMedicine Online* **16**(4): 504–513.

Tuttelmann F, Rajpert-De Meyts E, Nieschlag E and Simoni M (2007) Gene polymorphisms and male infertility – a meta-analysis and literature review. *Reproductive BioMedicine Online* **15**(6): 643–658.

CHAPTER 1

Genetic variation in key regulators of spermatogenesis in SSF: case - control studies in the Portuguese population

It is currently well established that a large number of genes involved in spermatogenesis are distributed across the genome, rather than restricted to the sex chromosomes. This is suggestive of a complex genetic architecture underlying spermatogenic failure and a subset of infertility phenotypes. In this chapter, I followed a candidate-gene approach to investigate the genetic variation in two key regulators of spermatogenesis. Diversity at these loci was interpreted considering the frequency in cases versus controls, alteration of the protein sequence, impact on transcriptional regulation and conservation status of the affected nucleotides/residues across species. This approach allowed the identification of new variants potentially associated with the most severe phenotype of male infertility: severe spermatogenic failure.

Introduction

Historically, mapping of a disease locus was initially performed by segregation and linkage analyses taking advantage of heritability data from families carrying the phenotype. Although a significant amount of genes have been associated with specific conditions using such approaches, these were mainly suitable for monogenic diseases showing a Mendelian-type of inheritance (Reviewed in Strachan and Read (2011a)). More complex multifactorial phenotypes revealed more challenging to tackle given that many environmental and genetic factors may collectively contribute to the disease. In this regard, genome-wide association studies (GWAS) using microarrays of common single nucleotide polymorphisms (SNPs) were extremely useful to start assessing the subset of genetic variants segregating in the population that were more frequent in cohorts of patients with phenotypes with more complex genetic architectures. (Reviewed Strachan and Read (2011b)). Recently, a study led by our team provided a significant contribution to unravel genetic variants overrepresented in patients with SSF (Lopes *et al.* 2013). In this work, by genotyping genome-wide single-nucleotide polymorphisms (SNPs) to detect copy-number variants (CNVs), the authors tested the hypothesis that infertile men carry more rare deleterious mutations than controls. At the time, this was the largest whole genome study investigating the role of rare variants in infertility, collecting data from 323 Caucasian infertile men and 1,136 controls. An independent sample of 979 Han Chinese men with idiopathic azoospermia and 1,734 controls as well as additional 4,519 controls from public databases were also assessed. The authors found a CNV burden in patients, where each rare autosomal deletion multiplicatively increased the risk of disease by 10%, rare X-linked CNVs by 29%, and rare Y-linked duplications by 88%. Importantly, this burden was unrelated to the disruption of haploinsufficient (HI) genes, since the estimated probability of deletion pathogenicity due to dominant disruption of a haploinsufficient gene (HI score) in infertility cases was indistinguishable from controls and much smaller when compared to cases of autism and developmental disorders. Although this suggests that overall SSF more likely results from large effect recessive mutations, or even the combined effect of deleterious mutations across many loci, two outlier genes were detected in the cohort of Portuguese patients in this analysis. The *Wilms' Tumor 1* (*WT1*) and the *Mitogen-activated protein kinase* 1 (*MAPK1*) genes, with the former computationally predicted to be strongly dosage sensitive. This signal resulted from a patient-specific deletion in the 11p13 genomic region (Seabra *et al.* 2014), known to be involved in gonadal development and differentiation (Barrionuevo *et al.* 2012). Also, a total of 5 deletions encompassing the *Doublesex and mab-3 related transcription factor 1* (*DMRT1*) gene were detected in the Utah and Han Chinese cohorts, that were absent from 6253 controls. DMRT1 is also a known regulator of testis differentiation and maintenance,

highly conserved in vertebrates (Reviewed in Zarkower (2013)). The combined results from this work propose DMRT1 loss-of-function mutations as a risk factor and potential genetic cause of human spermatogenic failure.

These results highlighted both *DMRT1* and *WT1* as relevant candidate genes for follow up studies. Given that with the array-based technique used in the previous study only common SNPs and large rare CNVs were genotyped, genetic screens that detect all the variation within the gene sequence, capturing small insertions/deletions (indels) and point mutations could potentially identify smaller sequence changes associated with male infertility in these loci.

Doublesex and mab-3 related transcription factor 1 - DMRT1

The DMRT protein family encloses a group of zinc-finger (ZF) transcription factors that bind the promoters of target genes through their highly conserved DM-domain. This DNA-binding motif was first described in the human genome (Raymond *et al.* 1998) and named after the structural similarity shared with the ZF domain identified in two invertebrate genes: the fly *doublesex* (*dsx*; Erdman and Burtis (1993)) and the nematode gene *male abnormal 3* (*mab-3*; Raymond *et al.* (1998)). Humans and mice contain 8 *DMRT* genes, 3 of which (DMRT1-DMRT3-DMRT2) cluster in a highly conserved vertebrate sexual development-associated locus located in the short arm of chromosome 9 [chr9:841690-1057554 (GRCh38/hg38 genome assembly); Reviewed in Bratuś (2012)]. Particularly, *DMRT1* has been extensively studied over the past decades in a plethora of vertebrate species, establishing this gene as a key regulator of the pathway of sex determination/gonad development in most species of this subphylum (Reviewed in Zarkower, 2013). Interestingly, it seems that the *DMRT1* gene was fine-tuned throughout evolution in response to clade-specific requirements: a) it is a switch-like sex-determining gene in reptiles employing a temperature-dependent sex-determining pathway; b) a dosage-sensitive gene on the Z chromosome that triggers testis development in birds; c) part of a sex-determining gene complex (*Dmrt1bY/DMY*) equivalent to a non-mammalian *sex-determining region Y* (*Sry*) in medaka fish; d) it has an essential role in male gonad differentiation but not sex determination in mice; and e) a partial role in sex determination [shared with *SRY*] and required for proper gonad development in humans (Reviewed in (Bratuś 2012)). Interestingly, while the main function of DMRT1 is to ensure proper development and maintenance of the male phenotype in the postnatal testis, overexpression of *Dmrt1* in XX mice

was sufficient for sex-reversal, suggesting that although not required, the function of DMRT1 in sex determination was maintained throughout evolution (Zhao *et al.* 2015).

DMRT1 spans a region of ~127 kb, comprising 5 exons (chr9:841,690-969,090; GRCh38/hg38 genome assembly). In humans, two additional exons, 6 and 7, are incorporated in isoforms DMRT1b and DMRT1c, respectively, through alternative splicing (Cheng 2006). Interestingly, *DMRT1* exhibits higher transcriptional diversity in other species. Four main isoforms generated by alternative splicing were described in mice, with 3 additional subtypes of the isoform Dmrt1a resulting from alternative polyadenylation signals in the 3'-UTR. These present similar expression patterns, being testis-specific in adult and more highly expressed in the male than the female gonad during embryonic development (Lu *et al.* 2007). In the Indian mugger, a species with temperature-dependent sex determination (TSD), eight isoforms are generated by alternative splicing with the exonization of intronic sequences and exon skipping and by alternative 3' polyadenylation (Anand *et al.* 2008). Their expression is significantly upregulated in male embryos from the start of the temperature sensitive period, suggesting a role for Dmrt1 transcripts in TSD. Interestingly, it appears that the plasticity of DMRT1 regulation in terms of gene expression and transcriptional diversity is influenced by transposable elements (TE). As previously stated, a duplicated copy of *Dmrt1* (*dmrt1bY* or *DMY*) is the equivalent of the mammalian *SRY* in sex determination in medaka fish. Recently, it has been shown that the insertion of a TE within its promoter region allowed for its transcriptional regulation by autosomal Dmrt1 (Herpin *et al.* 2010). Moreover, the exonization of Alu elements in alternative human *DMRT1* transcripts together with a non-random distribution of such elements across *DMRT1* introns suggests that they may have played a role in shaping the evolution of this genomic locus (Cheng *et al.* 2006).

Mouse models with loss of gene function (knock-out; KO) in all cells or in specific cell types/tissues (conditional KOs) were fundamental to better understand the role of *DMRT1* in gonadal development and maintenance. In mice, fetal *Dmrt1* is expressed in the genital ridges of both sexes, becoming male-specific restricted to Sertoli (SC) and germ cells (GC) after the activation of the *Sry* gene (Smith *et al.*, 1999; De Grandi *et al.*, 2000). In accordance, homozygous *Dmrt1* KO mice (*Dmrt1*$^{-/-}$) present severely hypoplastic testes with disorganized seminiferous tubules depleted of GCs, which further implicates *Dmrt1* in postnatal testis differentiation in a recessive manner (Raymond, 2000). In the juvenile murine testis, *Dmrt1* conditional KOs in Sertoli (SC*Dmrt1*KO) or germ cells (GC*Dmrt1*KO) have shown that DMRT1 production in SC is required for their postnatal differentiation and for germ line maintenance

and meiotic progression, whereas GCDmrt1 is needed for proper radial migration to form the spermatogonial stem cell niche, for mitotic reactivation and for general germ cell survival beyond the first postnatal week (Kim *et al.* 2007). These observations suggest that *Dmrt1* is required for proper gonadal development supporting cell function and for the establishment of undifferentiated spermatogonia in juvenile testis. Later, in the adult testis, absence of DMRT1 is the molecular switch that drives spermatogonial differentiation towards meiotic progression. GCDmrt1 represses *Stimulated By Retinoic Acid 8* (*Stra8*) and the Retinoic Acid (RA) pathway, which maintains spermatogonia in proliferation and differentiation preventing meiotic entry (Reviewed in (Zarkower, 2013)). Interestingly, the opposite role has been described for Dmrt1 in female fetal GCs, where it is needed for the transcriptional activation of *Stra8*. It appears therefore that *Stra8* is controlled sex-specifically by murine Dmrt1: activated in the fetal ovary and repressed by it in the adult testis (Krentz *et al.*, 2011). Importantly, Dmrt1 is fundamental in adult SCs to maintain the male phenotype by repressing the expression of *Forkhead box L2* (*Foxl2*), which avoids female reprogramming of the testis (Matson *et al.*, 2011). These studies demonstrate that DMRT1 plays a critical role in the postnatal development and maintenance of gametogenesis and testis architecture. Indeed, a breakthrough study by Murphy and co-workers has shown that DMRT1 binds to the promoter-proximal region of at least 1400 genes in the juvenile testis, and it is a bi-functional transcription factor activating some genes and repressing others (Murphy *et al.* 2010). Also, differential gene expression in response to the loss of DMRT1 in KO mice suggested that gene-specific and cell-type specific function of DMRT1 is likely modulated by other transacting factors. Importantly, the authors reveled the complexity of the molecular network directly regulated by DMRT1, comprising genes required for differentiation of Sertoli and germ cells, cell-cycle regulation, tight-junction dynamics and pluripotency.

In humans, data from the Human Protein Atlas (http://www.proteinatlas.org/%20ENSG00000137090-DMRT1/tissue) and the Genotype-Tissue Expression (http://www.gtexportal.org/home/gene/DMRT1) projects indicates that DMRT1 expression is similarly restricted to the adult testis. Specifically, DMRT1 is present in the nucleus of SCs and spermatogonia (Jorgensen *et al.* 2012). However, and contrary to the mouse model, *DMRT1* was found to co-express with *SRY* in the genital ridge of the male but not of the female embryo, supporting a role for DMRT1 in sex determination in humans (Moniot *et al.* 2000). Remarkably, a recent GWAS study identified *DMRT1* as a sex-specific genetic determinant of childhood-onset asthma (Schieck *et al.* 2016). In this study, the authors also detected an overexpression of DMRT1 in the cytoplasm of alveolar macrophages of patients

with interstitial lung disease and/or chronic obstructive pulmonary disease, not seen in normal tissue. This suggests a new role for DMRT1 in the context of disease, so far only linked with gonadal disorders and testicular tumorigenesis. In this respect, several syndromic human phenotypes have been associated with the hemizygosity of the genomic region of the *DMRT* gene cluster: a) XY sex reversal (Veitia *et al.*, 1997); b) true hermaphroditism (Õunap *et al.*, 2004) and c) gonadal dysgenesis (Tannour-Louet *et al.*, 2010), supporting a model of haploinsufficiency for this gene in humans. Interestingly, it appears that mutations in *DMRT1* alone are better tolerated, being absent from patients with gonadal dysgenesis (Machado *et al.*, 2012) and present in milder non-syndromic phenotypes such as non-obstructive azoospermia (NOA; Lopes *et al.* (2013); Tewes *et al.* (2014)).

Wilms' Tumor 1 – WT1

WT1 is located on human chromosome 11p13 (chr11:32,387,776-32,435,535; GRCh38/hg38 genome assembly) and comprises 10 exons coding for a protein containing a C-terminal ZF DNA- and RNA binding domain (exons 7-10), and an N-terminal self-association domain (exon 1). Strikingly, *WT1* displays a variety of transcript isoforms, at least 24, that are generated by a combination of alternative translation start sites, alternative RNA splicing and RNA editing (Toska and Roberts 2014). Importantly, transcriptional regulation of this gene seems to have relevant biological significance. In mammals, the 4 predominant isoforms result from alternative splicing of exons 5 and 9 (Haber *et al.* 1991; Gessler *et al.* 1992). The resulting isoforms display the inclusion/exclusion of exon 5 (17 amino acids) and of the coding sequence for three amino acids (lysine, threonine and serine (KTS)) located between the third and fourth ZF domains. Adding to the molecular repertoire, *WT1* transcripts are translationally regulated [non-AUG and alternative AUG translation initiation sites (TSS)] and proteins are post-translationally modified by phosphorylation, ubiquitylation and sumoylation (Reviewed in Huff (2011)).

As a bi-functional transcription factor, WT1 regulates a panoply of target genes involved in development and growth, differentiation, cytoskeleton organization and cell adhesion, WNT [wingless-type MMTV (mouse mammary tumor virus) integration site] signaling, MAPK signaling, apoptosis and epigenetic regulation. This activity is modulated by dimerization with other proteins (heterodimerization) that function as DNA-binding transcription factors, transcriptional co-regulators and epigenetic, proteolytic or RNA metabolism and processing

factors (Reviewed in Toska and Roberts (2014)). Homodimerization of WT1 through the N-terminal region has also been reported. Although the physiological role of self-association is not yet understood, it explains the dominant-negative effect of WT1 mutants on WT1 wild-type protein in the context of disease (Holmes *et al.* 1997). WT1 also acts as a post-transcriptional regulator, playing a role in RNA metabolism and translation. It has been shown to bind to specific RNA sequences, to interact with RNA-binding proteins and splicing factors and to associate with actively translating polysomes. Interestingly, the ability of WT1 to bind DNA and RNA seems to be influenced and dependent on the presence of the KTS amino acid motif. Lack of the KTS insertion (-KTS) results in a stronger DNA bond and such WT1 isoforms act as transcriptional regulators. WT1+KTS isoforms can also perform that function but are mainly associated with post-transcriptional processes. The biological relevance of these two isoform types is further strengthened by the fact that they are conserved from zebrafish to humans, with only two variations described in fish and non-mammals (Reviewed in Toska and Roberts (2014)).

Reflecting its critical role in development, WT1 is required in a large number of human tissues, including the gonads, kidneys, spleen, liver, lungs, heart and arteries, smooth and skeletal muscle, brain and spinal cord (Parenti *et al.* 2015). In accordance, WT1-null mice are embryonic lethal due to failed development of the kidneys, gonads, heart, diaphragm, spleen and adrenal glands and neuronal precursor tissues of the sensory system (Toska and Roberts 2014). During embryogenesis, the two WT1 KTS isoforms have been shown to play distinct roles in gonadal development. WT1 is initially expressed during embryonic development in the urogenital ridge, where the WT1–KTS isoform stimulates *SF1* (*Steroidogenic Factor 1*; also known as *NR5A1* (*Nuclear receptor subfamily 5, group A, member 1*), an orphan nuclear receptor; Wilhelm and Englert (2002)] and together promote the differentiation of the gonadal ridge. In this bi-potential gonad, WT1-KTS activates the expression of *SRY* [*sex-determining region Y*; Hossain and Saunders (2001)]. Then, both SRY and SF1 upregulate *Sox9* [*SRY box 9*] that initiates testicular differentiation. This is achieved by the production of Anti-Müllerian Hormone (AMH), responsible for the regression of the female Müllerian ducts, which is stimulated by SOX9 and enhanced by WT1 and SF1 (Rey and Grinspon 2011). The role of WT1 (+KTS) in sex determination is further supported by the observations that mice lacking this isoform do not express Sox9 and Amh (Hammes *et al.* 2001). WT1 is also expressed in many adult tissues (http://www.proteinatlas.org/ENSG00000184937-WT1/tissue; http://www.gtexportal.org/home/gene/WT1), although restricted to SCs in the testes, kidney podocytes, granulosa cells in the ovaries, mesothelium and 1% of bone marrow cells (Toska

and Roberts 2014), where it is essential for tissue maintenance. Ubiquitous ablation of WT1 in young and adult mice resulted in rapid deterioration of multiple tissues, resulting in kidney failure, atrophy of the spleen and pancreas and compromised erythropoiesis (Chau *et al.* 2011). Importantly, depletion of WT1 in adult mice testis disrupted the blood–testis barrier (BTB) and compromised polarity and maintenance of SCs, leading to a major loss of germ cells in the seminiferous tubules, with a phenotype resembling that of patients with NOA (Wang *et al.* 2013).

WT1 was named Wilms' tumor since it was first described to carry mutations causing a form of kidney cancer with the highest incidence in children (Eggers and Sinclair 2012). Three main clinical phenotypes are associated with constitutional WT1 mutations (Reviewed in (Seabra 2012)): 1) Large 11p13 deletions in WAGR syndrome (Wilm's tumor, aniridia, genitourinary anomalies and mental retardation); 2) Point mutations on the coding sequence of the ZF domain of WT1 in the Denish-Drash syndrome (Mesangial sclerosis, genital abnormalities, Wilms' tumor); and 3) intronic mutations in the second donor splice site of intron 9 in patients with Frasier Syndrome (XY pseudohermaphroditism and glomerulonephropathy). Other non-syndromic phenotypes can result from alterations in WT1 coding sequence, such as cryptorchidism, penoscrotal hypospadias, hypoplastic testes and renal disorders such as nephropathy or Wilms' tumor (Kohler *et al.* 2011). Indeed, the variable phenotypic expression of WT1 defects is related to alterations in regions coding for different protein domains. While mutations affecting the C-terminus ZF DNA-binding domain result in more severe gonadal dysgenesis (Little and Wells 1997; Royer-Pokora *et al.* 2004; Huff 2011), a milder phenotypic impact is expected for N-terminal variants, which are the most frequent type of variation found in NOA patients (Wang *et al.* 2013).

In summary (Fig. 3; Reviewed in (Eggers and Sinclair 2012)), WT1 is required for proper development of the bi-potential gonad and for sex determination by activation of SRY in the developing testis. At this stage, at least one copy of DMRT1 is required for germ cell survival. In the adult testis, WT1 and DMRT1 production in SCs is needed for germ cell survival and preservation of the male phenotype, respectively, whereas DMRT1 expressed by GCs maintains germ cell proliferation and differentiation by repression of Stra8 and the RA pathway. For both genes, hemizygosity at these loci has been reported as the cause of several gonadal

malformations and impaired fertility. Importantly, genetic variation in these genes appears to lead to a broad spectrum of phenotypic outcomes, as it has also been linked to milder defects of germ cell depletion in otherwise normal gonads. This may result from the disruption of specific sequence motifs either utilized by these proteins for their regulatory functions or recognized by proteins that control gene expression or transcript processing of *WT1* and *DMRT1* that could compromise specific signaling pathways involved in spermatogenesis maintenance but not gonadal development.

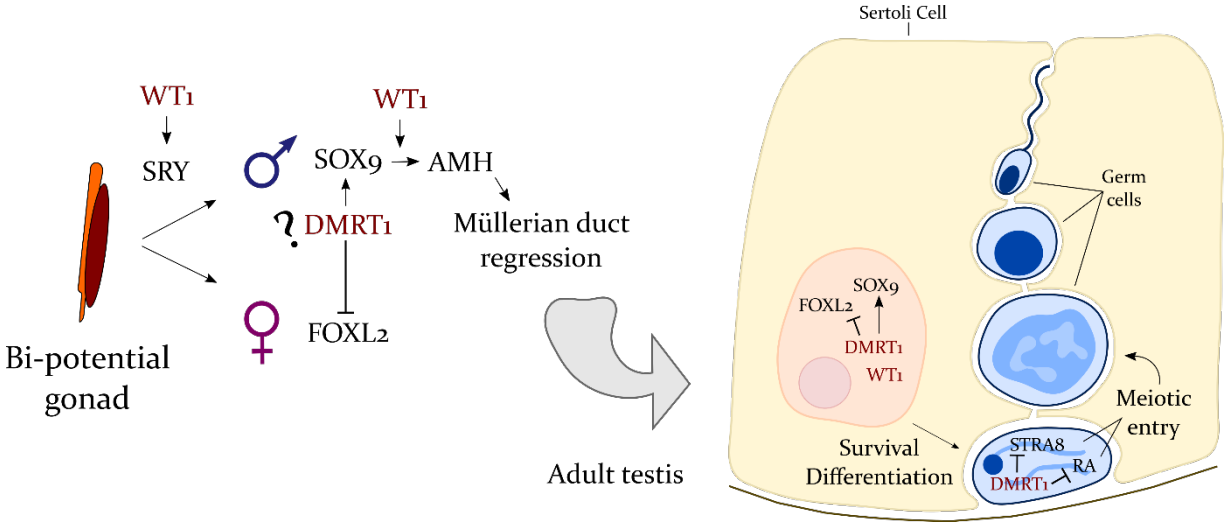


Figure 3. Regulation of gonadal development and maintenance by DMRT1 and WT1.

During embryogenic development, WT1 initiates and, together with other factors, maintains the male determining pathway. It is thus crucial for sex determination and normal testis development. In the fetal testis, after sex determination, DMRT1 is involved in germ cell fate possibly by a dual function of stimulating male factors while repressing the female counterparts. In the adult testis, WT1 expression is restricted to Sertoli cells and influences germ cell survival. DMRT1 in these cells plays a crucial and direct role in maintaining the male phenotype. It exerts positive regulation on male factors, through the stimulation of SOX9, and inhibits the female pathway by repression of FOXL2. In germ cells, DMRT1 maintains spermatogonia in proliferation and prevents differentiation by avoiding meiotic entry via Retinoic Acid (RA) pathway and Stra8 repression. This figure is meant to summarize the specific roles of WT1 and DMRT1 in the fetal and adult testis and not to comprehensively illustrate all the factors involved in these regulatory networks. Adapted from (Calvel *et al.* 2010; Matson *et al.* 2011; Eggers and Sinclair 2012).

Specific Objectives & Results

1- Genetic screening in patients with idiopathic SSF

It is well documented that male infertility phenotypes can arise from genetic alterations in a panoply of genes spread throughout the genome. While defects in some genes have large effects and may even result in major defects of urogenital development, there is a growing hypothesis that a combinatory effect of rare variants may unbalance the regulatory network controlling spermatogenesis progression and lead to spermatogenic impairment in the absence of other abnormalities. The *DMRT1* and *WT1* genes have been associated with both severe phenotypes with gonadal malformation, as well as SFF without urogenital defects. The main goal of the work described in this chapter was to search potential genetic causes of SFF within the genomic sequence of these candidate genes which showed large deletions in azoospermic males in our previous case-control genome-wide association study. These genes code for transcription factors known to play important roles in gonadal development and spermatogenesis maintenance and thus we hypothesized that other deleterious variants may be present in azoospermic patients. Genetic screens were conducted using a combination of different techniques, including Sanger sequencing, to detect base substitutions and small insertions/deletions within the coding and flanking intronic gene sequences. The frequency of the genetic variants found was compared between cases and controls and the impact on the protein or on the transcriptional regulation of the gene was predicted. Finally the conservation status of the affected nucleotides/residues across species was also assessed.

Paper II – Lima, A.C., Carvalho, F., Goncalves, J., Fernandes, S., Marques, P.I., Sousa, M., Barros, A., Seixas, S., Amorim, A., Conrad, D.F., Lopes, A.M. (2015). Rare double sex and mab-3-related transcription factor 1 regulatory variants in severe spermatogenic failure. *Andrology* **3**: 825-833.

Paper III – Seabra, C.M., Quental, S., Lima, A.C., Carvalho, F., Goncalves, J., Fernandes, S., Pereira, I., Silva, J., Marques, P.I., Sousa, M., Barros, A., Seixas, S., Amorim, A., Lopes, A.M. (2015) The mutational spectrum of *WT1* in male infertility. *The Journal of urology* **193**: 1709-1715.

[Note: Part of the work described in Paper III was performed under the scope of the Master's dissertation of C.M. Seabra (http://hdl.handle.net/10773/9537). My main contribution to this work concerns the genetic screening and analysis of the cohort of severe oligozoospermic patients and controls.]

Paper II - Rare double sex and mab-3-related transcription factor 1 regulatory variants in severe spermatogenic failure

Andrology, 2015

ANDROLOGY

ISSN: 2047-2919

ANDROLOGY

ORIGINAL ARTICLE

Correspondence:
Alexandra M. Lopes, IPATIMUP, Rua Dr. Roberto Frias s/n, 4200-465 Porto, Portugal.
E-mail: alopes@ipatimup.pt

Keywords:
alternative splicing, cis-regulatory variants, double sex and mab-3-related transcription factor 1, double sex and mab-3-related transcription factor 1 promoter, non-obstructive azoospermia

Received: 3-Nov-2014
Revised: 9-May-2015
Accepted: 11-May-2015

doi: 10.1111/andr.12063

Rare double sex and mab-3-related transcription factor 1 regulatory variants in severe spermatogenic failure

[1,2,3,4]A. C. Lima, [5]F. Carvalho, [6]J. Gonçalves, [5]S. Fernandes, [2,3,7]P. I. Marques, [8]M. Sousa, [5]A. Barros, [2,3]S. Seixas, [2,3,9]A. Amorim, [4,10]D. F. Conrad and [2,3]A. M. Lopes

[1]Graduate Program in Areas of Basic and Applied Biology (GABBA), Abel Salazar Institute of Biomedical Sciences, University of Porto, [2]Instituto de Investigação e Inovação em Saúde, Universidade do Porto, [3]Instituto de Patologia e Imunologia Molecular da Universidade do Porto, Porto, Portugal, [4]Department of Genetics, Washington University School of Medicine, St. Louis, MO, USA, [5]Department of Genetics, Faculty of Medicine of the University of Porto, Porto, [6]Department of Human Genetics, National Institute of Health Dr. Ricardo Jorge, Lisboa, [7]Abel Salazar Institute of Biomedical Sciences, University of Porto, [8]Laboratory of Cell Biology, UMIB, ICBAS, University of Porto, [9]Faculty of Sciences of the University of Porto, Porto, Portugal, and [10]Department of Pathology and Immunology, Washington University School of Medicine, St. Louis, MO, USA

SUMMARY

The double sex and mab-3-related transcription factor 1 (*DMRT1*) gene has long been linked to sex-determining pathways across vertebrates and is known to play an essential role in gonadal development and maintenance of spermatogenesis in mice. In humans, the genomic region harboring the *DMRT* gene cluster has been implicated in disorders of sex development and recently *DMRT1* deletions were shown to be associated with non-obstructive azoospermia (NOA). In this work, we have employed different methods to screen a cohort of Portuguese NOA patients for *DMRT1* exonic insertions and deletions [by multiplex ligation probe assay (MLPA); $n = 68$] and point mutations (by Sanger sequencing; $n = 155$). We have found three novel patient-specific non-coding variants in heterozygosity that were absent from 357 geographically matched controls. One of these is a complex variant with a putative regulatory role (c.-223_-219CGAAA>T), located in the promoter region within a conserved sequence involved in *Dmrt1* repression. Moreover, while DMRT1 domains are highly conserved across vertebrates and show reduced levels of diversity in human populations, two rare synonymous substitutions (rs376518776 and rs34946058) and two rare non-coding variants that potentially affect *DMRT1* expression and splicing (rs144122237 and rs200423545) were overrepresented in patients when compared with 376 Portuguese controls (301 fertile and 75 normozoospermic). Overall our previous and present results suggest a role of changes in DMRT1 dosage in NOA potentially also through a process of gene misregulation, even though *DMRT1* deleterious variants seem to be rare.

INTRODUCTION

Male infertility is a complex phenotype often arising from the inability of men to produce viable spermatozoa capable of fertilization. Non-obstructive azoospermia (NOA) is a severe form of male infertility characterized by the lack of spermatozoa in the ejaculate, which results from conditions other than hypothalamic-pituitary disease or obstruction of the male genital tract (Jungwirth *et al.*, 2013). It is estimated that NOA accounts for 10–15% of the cases of male infertility, affecting 1% of male individuals worldwide (Practice Committee of American Society for Reproductive Medicine in collaboration with Society for Male Reproduction and Urology, 2008). Importantly, over the past years several studies have been showing that a significant portion of the genes involved in severe spermatogenic failure (SSF) are spread throughout the genome rather than restricted to the sex chromosomes (reviewed in Lima & Lopes, 2014). One particular genomic region (9p), harboring the double sex and mab-3-related transcription factor (DMRT) gene cluster, has been linked with disorders of sex development (Veitia *et al.*, 1997; Raymond *et al.*, 1999; Õunap *et al.*, 2004; Tannour-Louet *et al.*, 2010). It appears that isolated genetic defects in *DMRT1*, one of the genes in the cluster, may be the cause of alterations in gonadal development in some cases (Ledig *et al.*, 2010, 2012) but not all, as no inactivating mutations or deletions were found in this gene in a study of patients with gonadal dysgenesis (Machado *et al.*, 2012). However, we have recently identified

recurrent 9p24.3 deletions spanning *DMRT1* in patients with idiopathic azoospermia, suggesting that loss of function of this gene is a risk factor and potential genetic cause of human SSF (Lopes *et al.*, 2013).

Double sex and mab-3-related transcription factor proteins are transcription factors that bind to the promoters of target genes through their zinc-finger DM-domain. This DNA binding motif is highly conserved and requires dimerization for recognition of a target consensus palindromic sequence (Erdman *et al.*, 1996). Also, in association with the DM domain, the DMRT protein family shares a less conserved DMRT1-like domain of yet uncharacterized function. Particularly, DMRT1 has been extensively studied in vertebrates and is known to be crucial in the pathway of sex determination and gonadal development in most, if not all, species of this group (reviewed in Zarkower, 2013). Indeed DMRT1 has acquired specialized functions in different clades. It is essential for male gonad differentiation but not sex determination in mice, showing testis-specific expression restricted to Sertoli and pre-meiotic germ cells. A similar pattern is observed for human *DMRT1* with the exception that it is also part of the network of sex determination (reviewed in Bratus, 2012), being co-expressed with sex-determining region (SRY) in the genital ridge of the male embryo (Moniot *et al.*, 2000). Interestingly, a total of three cDNA isoforms have been detected in the human testis – DMRT1a, DMRT1b, DMRT1c – and result from the inclusion of intronic regions and *Alu* elements (Cheng *et al.*, 2006), suggesting that the transcriptional diversity of this gene might be involved in the regulation of spermatogenesis.

Bearing in mind the expression pattern of *DMRT1*, its crucial role in the progression of spermatogenesis and previous findings of an association of *DMRT1* defects with SSF, we sought to evaluate the contribution of sequence variants in the promoter and coding regions of the *DMRT1* gene to NOA phenotypes. Here, we describe novel non-coding variants and rare non-coding and synonymous variants overrepresented in NOA patients when compared with Portuguese fertile and normozoospermic controls, which are predicted to affect the regulation of the canonical *DMRT1* transcript (DMRT1a) and alternative splicing.

MATERIALS AND METHODS

Patient and control populations

Deoxyribonucleic acid samples of 155 male individuals with idiopathic NOA were collected at the Genetics Department from INSA-IP and at the Genetics Department from FMUP. Patients with known causes of infertility, including chromosome anomalies and Yq microdeletions were excluded from this study. The majority of these samples (n = 110) had been analyzed for CNVs by SNP microarrays (Lopes *et al.*, 2013) but none presented CNVs within *DMRT1*.

As controls, we obtained DNA of 376 Portuguese men: 75 normozoospermic (normal sperm parameters) and 301 males who fathered at least one child.

Ancestry of one NOA patient harboring two potential regulatory variants was determined by genotyping of 46 ancestry-informative markers (insertions and deletions). Biogeographical ancestry was assigned using Snipper 2.0 (http://mathgene.usc.es/snipper/; see Data S1 and Figure S1).

The study was included in the project 'Copy number variation in infertile men genomic regions: screening in the Portuguese population' (PTDC/SAU-GMG/101229/2008), approved by the INSA Ethics Committee (Lisbon, Portugal on 6 November 2007). Molecular studies were performed after informed consent and all DNA samples were anonymized.

Analysis of DMRT1 sequence

Multiplex ligation probe assay

To detect deletions/duplications in the coding region, MLPA was performed in 68 randomly picked NOA patients. SALSA MLPA P334 Gonadal probemix (MRC-Holland, Amsterdam, The Netherlands), which contains two probes for each exon of the gene, was employed according to the manufacturer's instructions. Typically 150–200 ng of genomic DNA were used for amplification.

Sanger sequencing

Primers for amplification and sequencing (Table S1) of the *DMRT1* genomic reference sequence (NM_021951.2) were designed in Primer3 v.0.4.0 (Koressaar & Remm, 2007; Untergasser *et al.*, 2012). DNA fragments were amplified by standard PCR, sequenced with the BigDye Terminator v.3.1 Cycle Sequencing Kit (Applied Biosystems, Life Technologies, Carlsbad, CA, USA) and run on an ABI PRISM 3130xl Genetic Analyzer (Applied Biosystems, Life Technologies). Because of low DNA quality, in 23 patients only a partial gene sequence was obtained, which was taken into account in the variant frequency calculations, performed independently for each exon (see Data S1). All individuals showing novel/rare variants had the full *DMRT1* sequence analyzed. Sequences were assembled and analyzed using Geneious version 5.5.8 (Biomatters, Auckland, New Zealand) and all putative polymorphisms were manually inspected and individually confirmed.

Genotyping of Portuguese fertile and normozoospermic controls

Rare variants (<1%) in European populations which were overrepresented in patients and novel mutations identified exclusively in NOA cases were genotyped in our Portuguese controls (Tables 1 and 2). The presence of two indels (c.354 + 38_insG with a juxtaposed poly-T and rs59834456) was detected by fragment size separation in agarose and polyacrylamide gels, respectively; variants in the promoter, 5′-UTR and CDS of exon 1 were genotyped by Sanger sequencing; six other variants were analyzed in a single SNaPshot Multiplex reaction. All variants detected by SNaPshot (Figure S2) were confirmed by Sanger sequencing. See Data S1 and Tables S1–S3 for more details on experiments and primers used.

In silico and statistical analysis

Sequence data were retrieved from the 1000 genomes Project (Genomes Project Consortium *et al.* 2012), NHLBI GO Exome Sequencing Project (ESP) and CLINSEQ projects to check for all the reported variants in the regions screened and used to set a group of unphenotyped controls of European ancestry (Data S1 and Table S4).

Allele counts of rare variants that were overrepresented in patients were compared with the Portuguese control population

Table 1 Novel *DMRT1* non-coding variants restricted to patients

Name	Genomic location	Genomic context	Variant type	Minor allele	MAF NOA	Global MAF	MAF PT controls
c.-223_-219CGAAA>T	chr9:841616-841620	Promoter Region	Upstream gene variant	T	1/278	0	0/714
c.355-6T>C	chr9:846954	Intron 1	Intronic (Splice region)	C	1/280	0	0/700
c.823-64_823-62delATT	chr9:916699-916701	Intron 3	Intronic	−ATT	1/290	0	0/660

DMRT1, double sex and mab-3-related transcription factor 1; MAF, Minor allele frequency; NOA, non-obstructive azoospermia; PT, Portuguese.

Table 2 Annotated variants detected in the cohort of Portuguese NOA patients

Reference name	Genomic location	Genomic context	Variant type	Minor allele detected	Patients MAF[a]	Global MAF[b]	EUR MAF[c]	PT controls MAF[d]	*p*-value*
rs144122237	chr9:841785	5′-UTR	5′-UTR variant	T	0.0036	0.004	0.0000	0.0000	0.2875
rs3739584	chr9:841825	5′-UTR	5′-UTR variant	T	0.1475	0.141	0.1378	0.1362	–
rs3739583	chr9:841971	Exon 1	Missense	A	0.1475	0.132	0.1227	0.1344	–
rs376518776	chr9:842051	Exon 1	Synonymous	A	0.0036	0.000	0.0006	0.0000	0.3440
rs55905583	chr9:846952	Intron 1	Intronic (Splice Region)	C	0.0071	0.011	0.0006	0.0057	1.0000
rs2273929	chr9:847208	Intron 2	Intronic	G	0.4607	0.417	0.5409	–	–
rs16925431	chr9:894036	Exon 3	Synonymous	C	0.0109	0.032	0.0222	–	–
rs146975077	chr9:894147	Exon 3	Synonymous	C	0.0036	0.009	0.0002	0.0014	0.4866
rs34946058	chr9:894156	Exon 3	Synonymous	G	0.0109	0.042	0.0007	0.0043	0.3596
rs59168737	chr9:894224	Intron 3	Intronic	A	0.0109	0.039	0.0220	–	–
rs1033836	chr9:894297	Intron 3	Intronic	A	0.0471	0.101	0.0409	–	–
rs59834456	chr9:894336–894339	Intron 3	Intronic	−ATAT	0.0109	NA	NA	0.098**	–
rs112866575	chr9:967939–967940	Intron 4	Intronic	+CTCCCTTT	0.0229	0.149	0.0985	–	–
rs200423545	chr9:967959	Intron 4	Intronic	A	0.0099	0.002	0.0024	0.0057	0.4291
rs376062302	chr9:967953–967954	Intron 4	Intronic	+TTCTCTCT	0.0915	0.1236[e]	NA	–	–
rs279895	chr9:967981	Intron 4	Intronic (Splice Region)	G	0.1765	0.111	0.1658	–	–
rs79358387	chr9:968150–968151	3′-UTR	3′-UTR variant	+T	0.1144	0.223	0.1135	–	–
rs279894	chr9:968334	3′-UTR	3′-UTR variant	G	0.1800	0.077	0.1541	–	–

rs59834456 was detected in two Bantu individuals but no population data are available in the Ensembl release of March 2015. NA, not available; NOA, non-obstructive azoospermia; MAF, Minor allele frequency detected in: [a]The NOA Portuguese cohort. [b]All populations screened in large genome sequencing projects. [c]European populations of the 1000 Genomes (EUR and CEU), ESP (European-American), and CLINSEQ (CSAgilent). [d]The Portuguese (PT) control population. [e]Control population used by Machado *et al.*, 2012. *Fisher's exact test using only the PT population as control. **Frequency of carriers.

applying Fisher's exact test, using RStudio v.0.98.1080 (R Development Core Team 2012).

Also, we predicted in silico the effect of novel non-coding variants in mRNA splicing [Human Splicing Finder v.2.4.1 (Desmet *et al.*, 2009); BDGP: Splice Site Prediction by Neural Network, available at http://www.fruitfly.org/seq_tools/splice.html and the transcription factor binding sites (TFBSs) within the promoter [MatInspector, included in the Genomatix Software suite (Genomatix Software GmbH, Munich, Germany)]; TFSEARCH 1.3, available at http://www.cbrc.jp/research/db/TFSEARCH.html.

We have retrieved the protein sequences of 12 vertebrates and *C. elegans* from Ensembl database of *DMRT1* orthologs. Manually curated protein sequences were then aligned using ClustalW.

RESULTS

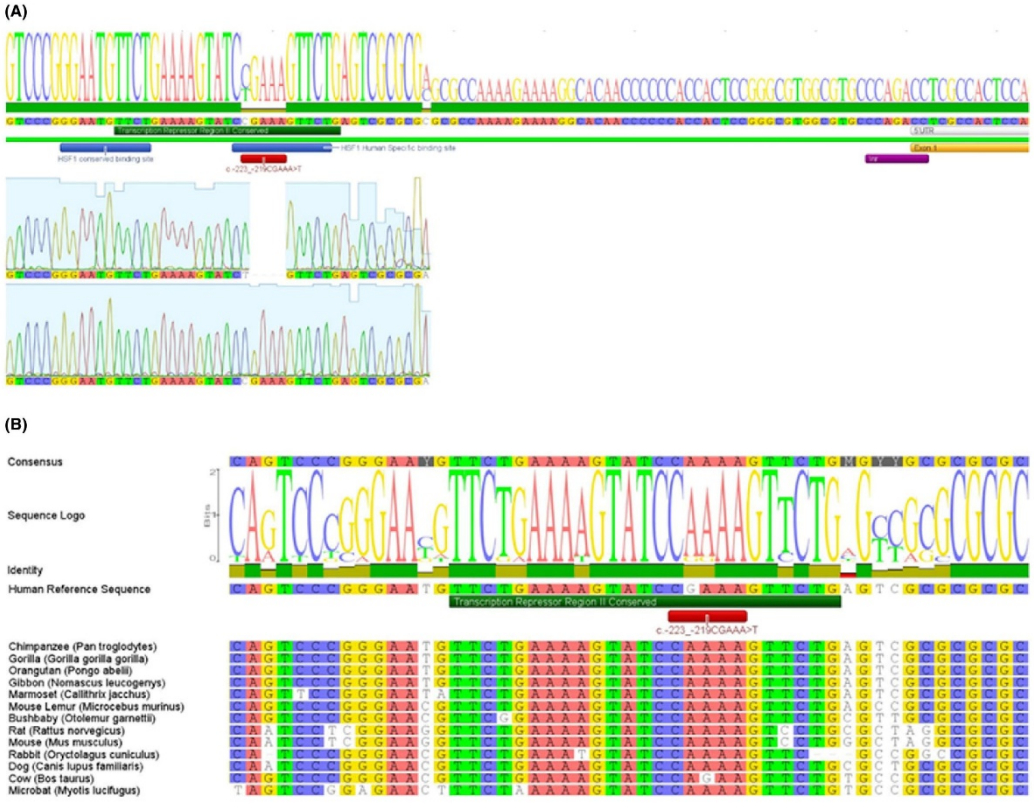
To characterize the mutational spectrum of potentially deleterious *DMRT1* variants in men with idiopathic SSF, we performed a survey of intragenic deletions/duplications by MLPA and point mutations by Sanger sequencing. By MLPA we aimed to find intragenic *DMRT1* rearrangements that might be more frequent than the large CNVs previously described in American and Chinese NOA patients (Lopes *et al.*, 2013). No alterations were detected by MLPA after screening approximately half of our patient cohort (n = 68) and therefore we conclude that, if present, intragenic *DMRT1* rearrangements are also rare in our population.

A new *DMRT1* promoter variant restricted to NOA

Given the potential deleterious effect of base substitutions and smaller deletions/duplications, we have screened the *DMRT1* sequence of 155 NOA patients by Sanger sequencing. In total we detected four novel non-coding variants in heterozygosity in patients (Table S2), three of which were absent from Portuguese controls (Table 1). All novel variants were further analyzed in silico for a regulatory role (see Materials and Methods and Data S1) and those with a predicted effect are described.

Interestingly, one of the variants detected only in a single NOA patient presents a complex genotypic configuration and is located within a region of the *DMRT1* known to be involved in the transcriptional repression of the gene in mice (Lei & Heckert, 2002). We identified, by allele-specific amplification followed by Sanger sequencing, a single variant (c.-223_-219CGAAA>T; Table 1) in one haplotype (i.e. inherited from the same parent, Fig. 1A) that most likely resulted from a complex mutation event. Importantly, this variant was not present in a sample of

Figure 1 Double sex and mab-3-related transcription factor 1 (*DMRT1*) promoter sequence in the proximal region of the transcription start site (TSS). The *DMRT1* promoter region in humans (A) harbors an Inr element (purple filled box), which is characteristic of TATA-less genes, and also one region associated with DMRT1 repression from −88 to −63 bp upstream the TSS (green filled boxes). We have found by allele-specific amplification a complex variant in the promoter of one azoospermic patient (A; red filled boxes). Analysis of the proximal region of TSS has predicted two binding sites for Heat-shock factors 1 (HSF1) within the conserved motif (A; blue boxes), one of which is disrupted by this novel patient-specific variant. Aligning the human sequence of this motif with the sequences from several species representative of mammalian clades (B) shows a high degree of conservation for this region (green filled boxes). Sequence annotations were performed using Geneious v.5.5.8 Software and alignments were obtained applying the ClustalW algorithm available in the Geneious v.5.5.8 Software.



357 geographically matched control individuals (301 fertile and 56 normozoospermic men). To assess its potential impact, we performed a more refined in silico analysis of the human *DMRT1* promoter (HG_KWN:62427), located within a CpG-island (chr9:840690–842192; NCBI37/hg19), a feature often associated with TATA-less promoters. Typically the core promoter activity and basal gene transcription of TATA-less genes is regulated by an initiator element (Inr), with a consensus sequence Py Py A + 1N T/A Py Py (Smale & Baltimore, 1989), which we have found in the promoter region of *DMRT1* (Fig. 1A) within the canonical transcription start site.

In the alignment of the *DMRT1* promoter sequences from several species representative of different mammalian clades (Fig. 1B), a high degree of conservation is apparent, in accordance with previous findings (Lei & Heckert, 2002). Moreover, no variants have been described in this region in human populations (chr9:841602–841626; GRCh37/hg19 assembly). In addition, we found that the G nucleotide (chr9:841617; hg19 assembly), deleted in the patient presenting the c.-223_-219CGAAA>T variant, is only found in the human sequence, whereas the ancestral allele A is present in all the other species. A screening for TFBSs within this conserved region of the *DMRT1* promoter was performed using two bioinformatics tools: MatInspector and TFSEARCH 1.3 (see Materials and Methods and Table S5). In the human reference sequence, binding of Heat-shock factors 1 (HSF1; +strand) and 2 (HSF2; −strand) was predicted by both programs. When the human-specific G is replaced by the ancestral A at this position or by the sequence variant found in the patient, a disruption of the HSF1 binding site (Fig. 1A) is predicted, whereas the score for HSF2 binding remains unchanged. Moreover, in the proximal region of the *DMRT1* transcription start site, the human reference sequence

contains a second binding site for HSF1 (HSF and TFSEARCH 1.3 predictions) that is conserved between human (chr9:841,690–969,090; GRCh37/hg19 assembly) and mouse (chr19:25,505,706–25,604,328; GRCm38/mm10 assembly).

Another patient was heterozygous for the insertion of a G nucleotide (c.354 + 38_insG) and for repeat length in a juxtaposed poly-T (>10 T) in intron 1. While this poly-T region seems to have small length variability in both Portuguese patients and controls, it is ≥10 bp longer than the reference sequence on both chromosomes of this patient. Length variability in this region alters the sequence of a *DMRT1* alternative transcript (DMRT1c) that has been previously detected in human testis (Cheng *et al.*, 2006) and may lead to a frameshift in the corresponding protein isoform depending on the number of repeats.

Rare *DMRT1* variants may play a role in splicing regulation

A subset of variants identified in the NOA cohort (n = 18) had been previously found in unphenotyped individuals from large genome sequencing projects (Table 2). We obtained the frequency of all *DMRT1* variants detected and set a control group of populations of European ancestry (see Data S1). As any genetic variant leading to NOA should be rare in the population, we have selected those at ≤1% in European populations and overrepresented in patients for genotyping in the Portuguese controls (Table 2; rs3739584 and rs3739583 were genotyped in the same assay as rs144122237 and rs376518776). While only the rs59834456 deletion was present in Portuguese normozoospermic controls, two variants (one synonymous substitution and a 5′-UTR variant) were also absent from fertile individuals with unknown sperm counts (Table S2). The synonymous substitution, rs376518776 (c.213G>A), juxtaposed to the DNA binding DM domain, was found in heterozygosity in a single NOA patient. It showed a six-fold increase in frequency in the NOA Portuguese cohort when compared with unphenotyped database controls of European ancestry and was absent from fertile and normozoospermic Portuguese controls. In silico predictions using the Human Splicing Finder Software (HSF; Desmet *et al.*, 2009) suggest that this substitution creates a new branch point site, a cis-element required for spliceosome assembly, with a score (89.44) above the average of naturally occurring branch point sequences (83.4 ± 8.6; Desmet *et al.*, 2009). This new branch point site may interfere with the interplay of splicing elements in its vicinity and potentiate the occurrence of new alternative transcripts.

The single rare 5′-UTR variant (rs144122237; c.-54C>T) resides within a putative binding site for the transcription factor RFX3 (Regulatory factor X, 3; Table S6) and was found in one NOA patient (0.34%) but was absent from Portuguese fertile and normozoospermic controls and from European populations. Given the presence of the same allele in Africans at low frequency (1.8%) we genotyped a panel of 46 ancestry-informative markers (Pereira *et al.*, 2012), and assigned an African ancestry for this patient (LR African vs. European = 7.712×10^{22}; see Data S1). This was the same individual who harbored two longer poly-T alleles (>10 bp) within intron 1, which may reflect his African genetic background.

Other coding variants overrepresented in patients are rs146975077 (c.774G>C) and rs34946058 (c.783C>G), the latter possibly interfering with splicing. In fact, using the HSF software two novel splicing enhancer motifs were predicted for the binding of SC35 (score 82.18) and SRp40 (score 79.28), both with scores above the respective thresholds (75.05 and 78.08). These proteins are members of the SR protein family known to be involved in the regulation of alternative splicing (Liu *et al.*, 1998). Nonetheless, this substitution was present in two Portuguese fertile controls, one of them homozygous for the minor allele.

Three rare (≤1%) intronic variants found in patients were also detected in unphenotyped individuals of large genome sequencing projects and are predicted to affect splicing: rs55905583 (c.355-8G>C) increases the likelihood of expression of the canonical isoform, rs59834456 (c.822 + 141_822 + 144delATAT) creates a new enhancer site for SRp40 (score 91.86) and rs200423545 (c.968-26T>A) creates a new branch point in the vicinity of exon 5 (HSF score of 95.11). While rs200423545 is present at nearly twice the frequency in patients compared with Portuguese controls, rs55905583 and rs59834456 are found at similar frequencies in the Portuguese controls (Table 2).

DMRT1 domains are highly conserved in mammals and show lower genetic variation in controls

Even though DMRT1 is known to be involved in the sex determination pathway from *C. elegans* to humans, there is much uncertainty regarding the role of each functional domain. To grasp the potential functional impact of *DMRT1* variants, we analyzed the spectrum of coding variants found in unphenotyped control populations (Table S7). In Fig. 2, we show the alignment of the available vertebrate DMRT1 sequences, and the invertebrate *C. elegans*, with variants found in human populations marked along the sequence. The zinc finger-like DM domain is involved in binding to the promoter of its target genes and therefore missense variants in this region are expected to have a functional impact. Indeed, only one missense variant (rs201947617) is described in the DM domain, in a less conserved region of the sequence. The DMRT1-like domain, typically found in association with the DM domain but of unknown function, is also conserved in mammals, even though some sequence variability can be found in the central region. Missense mutations within the conserved region are predicted to be deleterious, with an average frequency in the populations surveyed (0.0307%) similar to those without predicted functional impact on the protein (missense tolerated, 0.0266%; and synonymous variants, 0.0225%). Outside the domains, in regions not predicted to affect protein function, missense as well as synonymous substitutions are rare but more frequent (1.75% and 1.23%, respectively) than potentially damaging missense variants (0.026%), which present a similar frequency within the domains (0.028%).

Compared with the rest of the *DMRT1* coding sequence (67/729 = 0.0919) domains also show less variants per base pair (DM: 3/141 = 0.0212; DMRT1-like: 17/252 = 0.0675), likely reflecting functional constraints. Also, the high conservation of a stretch of 83 amino acids comprising the DM and part of the DMRT1-like domains within vertebrates supports their importance for protein function (Fig. 2A). Another region (19 amino acids) in the vicinity of DMRT1-like domain stands out in the alignment as highly conserved in mammals, also exhibiting shared homology with avian DMRT1 (Figure S3).

Figure 2 Alignment of Double sex and mab-3-related transcription factor 1 (DMRT1) protein sequences from several species representative of vertebrate clades, including human. A high degree of identity is notably shared by vertebrates within the DM domain. Actually, a stretch of 83 amino acids containing the DM and part of the DMRT1-like domains seems to be fairly maintained among these clades. Interestingly, the DMRT1-like domain is greatly conserved in its extremities and more variability can be found at its center. This alignment was obtaining applying the ClustalW algorithm available in the Geneious v5.5.8 Software. Amino acids are colored when they match at least 75% of the sequences. Purple filled boxes mark the domains; yellow and red filled boxes represent missense variants annotated in the Ensembl database, respectively, predicted by at least one software (SIFT or PolyPhen) as not affecting the protein or as deleterious.

DISCUSSION

The role of DMRT1 orthologs in sex determination as well as in gonadal development and maintenance throughout evolution is well established in the literature. In fact, DMRT1 is crucial in the maintenance of the mammalian gonad and required for normal progression of spermatogenesis. Expression of *Dmrt1* is necessary in Sertoli cells to prevent female reprogramming of the adult testis (Matson *et al.*, 2011) and in spermatogonia to maintain their proliferative state by avoiding meiotic entry (Zarkower, 2013). Recently, Murphy and coworkers have shown that DMRT1 is a bifunctional transcriptional regulator, activating genes required for proper development of the male phenotype while repressing others that are specific of the female gonad (Murphy *et al.*, 2010). These observations, together with the existence of human phenotypes associated with *DMRT1* alterations make this gene a strong candidate for a causal role in some of the cases of idiopathic NOA.

Following our initial discovery of recurrent *DMRT1* whole gene deletions in azoospermic patients from Utah (USA) and China

(Lopes *et al.*, 2013), we now describe novel and rare potentially regulatory variants in Portuguese patients, reinforcing that rare variants affecting *DMRT1* function may underlie some cases of severe spermatogenic impairment. In one NOA patient, we uncovered a non-coding variant in the promoter (c.-223_-219CGAAA>T) that resides within a region shown to be involved in the repression of *Dmrt1* expression in rat and mouse Sertoli cells (Lei & Heckert, 2002). Our in silico analysis not only supports the high conservation of this region across mammals (Fig. 1B) but also predicts that the human sequence is recognized by two heat-shock transcription factors (Fig. 1A and Table S5), HSF1 and HSF2. Heat-shock transcription factors are known to activate heat-shock response genes, under stress conditions. Interestingly, these two transcription factors (HSF1 and HSF2) are highly expressed in testis, showing germ cell type specificity and are required for the progression of spermatogenesis. In adult mice testis, HSF2 expression is limited to spermatocytes and spermatogonia (Kallio *et al.*, 2002), whereas HSF1 is highly and specifically expressed in the nuclei of spermatocytes and round spermatids (Akerfelt *et al.*, 2010). Apparently, *Hsf1* is specifically expressed in those stages of spermatogenesis where *Dmrt1* is not, suggesting that HSF1 might block the transcription of *Dmrt1*, therefore allowing spermatogenesis to progress.

Our analysis indicates that the base substitution within the repression motif of the human sequence has resulted in the acquisition of a second HSF1 binding site (Fig. 1 and Table S5) absent in the ancestral sequence (Fig. 1). Remarkably, this human-specific HSF1 binding site is disrupted by the c.-223_-219CGAAA>T variant, suggesting that it may result in *DMRT1* misregulation and the development of NOA. Indeed, the histology of this patient revealed maturation arrest (Table S2), resembling the phenotype of *Hsf1−/−* male mice which display regions of the seminiferous tubules containing only spermatogonia (Akerfelt *et al.*, 2010). It is noteworthy that a mutation in *HSF2*, also predicted to bind the *DMRT1* promoter, has been associated with azoospermia (Mou *et al.*, 2013). It would be interesting to further explore the role for HSF1 in *DMRT1* regulation during spermatogenesis to understand how its dynamics can contribute to the phenotype.

Three alternatively spliced *DMRT1* transcripts have been described in the human testis: DMRT1a, DMRT1b, and DMRT1c (Cheng *et al.*, 2006). DMRT1a is the canonical DMRT1 isoform, showing the highest expression, and translates into a 375 amino acid protein. Both DMRT1b and DMRT1c result from the incorporation of intronic sequences in the mature mRNA (intron 3 and intron 1, respectively) and are less expressed. Canonical splice signals are known to be insufficient for proper splicing, requiring the regulation provided by cis-elements such as enhancers or repressors, especially important when alternative splicing is involved (Cartegni *et al.*, 2003). Indeed, the creation of a new consensus exonic splicing enhancer (ESE) motif by a deep intronic mutation has already been shown to result in the production of an aberrant mRNA involved in afibrinogenemia (Davis *et al.*, 2009).

We have identified a novel indel that is predicted to alter splicing by interfering with the reading frame of DMRT1c (c.354 + 38_insG and linked poly-T in intron 1, Table S2). However, the length variation in the poly-T in intron 1 was also observed in our controls, making it a less likely candidate for a strong deleterious effect on DMRT1 function. Also, an African ancestry was determined for this NOA patient suggesting that this locus harbors longer alleles in African populations when compared with Europeans.

One rare (≤1%) non-coding variant more frequent in patients than in Portuguese fertile controls was also predicted to affect splicing (rs200423545, Table 2).

Furthermore, the single rare variant detected in the 5′-UTR (rs144122237) of one patient potentially disrupts a binding site for the transcription factor RFX3, which is highly expressed in the testis. This variant has not been detected in the European control populations surveyed (our own and those included in databases) and was only present in the patient with African ancestry. The relatively low frequency of this variant in Africans (1.8%; six female and three male carriers) does not preclude an effect on fertility and overall our results highlight the need to consider the combinatorial effect of regulatory variants, given the interplay between factors with different roles (expression regulation, splicing, etc.), as well as the genetic background of each individual.

Increasing data has been showing that synonymous variants might have an effect on transcription, splicing, mRNA transport or translation (Reviewed in Goymer, 2007). In accordance, we have found two synonymous variants overrepresented in patients that were predicted to potentially alter splicing by the creation of either new branch point sequences (rs376518776; c.213G>A) or new motifs for ESEs (rs34946058; c.783C>G). At the rs34946058 triallelic locus the T allele, and not the G allele found in our patients, has been recently reported in two men of German ancestry with cryptozoospermia (sperm concentration <0.1 million/mL), but also in one normozoospermic control (Tewes *et al.*, 2014). Interestingly, the Portuguese population seems to harbor only two alleles at this locus, the reference C and the minor allele G that reaches a frequency of approximately 6% in African populations. Three patients carried the G allele, all of which had been previously analyzed for genome-wide SNPs (Lopes *et al.*, 2013) and show no evidence for African ancestry/admixture. One fertile control also carried the G allele and another fertile control was homozygous for the same allele. Thus, even though the sperm parameters for the fertile G homozygous individual remain unknown, the relatively high frequency of this allele in Africans does not support a strong impact on fertility.

Overall, our in silico functional analysis suggests that some of the of *DMRT1* variants found in patients, which are not predicted to affect the protein (synonymous and non-coding) may, however, be involved in alternative splicing and other regulatory mechanisms. Per se, these variants are likely to have a mild impact on fertility, probably affecting the regulatory network controlling spermatogenesis but may also be in linkage disequilibrium with intronic cis-acting SNPs and therefore a combinatorial effect of regulatory variants could be implicated in NOA. Moreover, the lack of deleterious coding mutations and intragenic rearrangements (screened by MLPA) in our Portuguese cohort supports the view that *DMRT1* genetic defects are rare (Lopes *et al.*, 2013; Tewes *et al.*, 2014) and strong association for individual variants may only be detected in multicenter studies with larger cohorts.

Concordantly, our analysis of the DMRT1 protein among vertebrates (Fig. 2) reinforces the previous observations. The highly conserved DM domain as well as some regions of DMRT1-like

domain seem to be under stronger functional constraints displaying less variation. Together with our previous results for CNVs (Lopes *et al.*, 2013), and a recent screening for *DMRT1* variants in German patients (Tewes *et al.*, 2014), the finding of variants potentially affecting *DMRT1* expression supports a model where changes in dosage in this gene may underlie some cases of NOA. The association of *DMRT1* hemizygosity with human disorders of sex development (Veitia *et al.*, 1997; Raymond *et al.*, 1999; Ounap *et al.*, 2004; Barbaro *et al.*, 2009; Tannour-Louet *et al.*, 2010) but also NOA (Lopes *et al.*, 2013; Tewes *et al.*, 2014), further points to other cis- and trans-acting factors contributing to these more complex phenotypes. It seems therefore that both *DMRT1* copies must be functional for proper gonadal development and spermatogenesis maintenance and that variation affecting one or both alleles would result in phenotypes of increasing severity.

In summary, we provide evidence suggesting that rare regulatory variants in coding and non-coding regions of the *DMRT1* gene and its promoter might be involved in idiopathic NOA. Altogether our results draw attention to the potential contribution of regulatory variants to male infertility, namely NOA, reinforcing the complexity and significance of regulatory networks involved in the control of spermatogenesis.

ACKNOWLEDGMENTS

The authors thank Rui Pereira (Instituto de Patologia e Imunologia Molecular da Universidade do Porto, Porto, Portugal – IPATIMUP) for the analysis of ancestry performed for one patient. This work was partially funded by the Portuguese Foundation for Science and Technology FCT/MCTES (PIDDAC) and co-financed by European Funds (FEDER) through the COMPETE program (PTDC/SAU-GMG/101229/2008 to AML), CIGMH and Pest-OE/SAU/UI00009/2011. IPATIMUP is an Associate Laboratory of the Portuguese Ministry of Science, Technology, and Higher Education and is partially supported by FCT. AML and ACL are funded by FCT fellowships SFRH/BPD/73366/2010 and SFRH/BD/51695/2011, respectively. DFC was supported by grant number R01HD078641 from the Eunice Kennedy Shriver National Institute of Child Health and Development of the United States National Institutes of Health.

DISCLOSURE

The authors declare no competing interests.

AUTHOR CONTRIBUTIONS

ACL performed the *DMRT1* sequence analysis in Portuguese patients and controls, conducted the data analysis, extracted DNA of fertile control samples, interpreted the results, and drafted the manuscript. FC, JG, SF, MS, and AB performed the clinical and genetic characterization of the patients' samples for the routine workup of male infertility. PIM performed DNA extraction of the normozoospermic patients and contributed to the genotyping of the controls. AA, SS, and DFC contributed with a critical review of the manuscript. AML obtained financial support, designed the study and supervised the data analysis, interpretation of results, and manuscript preparation.

REFERENCES

Akerfelt M, Vihervaara A, Laiho A, Conter A, Christians ES, Sistonen L & Henriksson E. (2010) Heat shock transcription factor 1 localizes to sex chromatin during meiotic repression. *J Biol Chem* 285, 34469–34476.

Barbaro M, Balsamo A, Anderlid BM, Myhre AG, Gennari M, Nicoletti A, Pittalis MC, Oscarson M & Wedell A. (2009) Characterization of deletions at 9p affecting the candidate regions for sex reversal and deletion 9p syndrome by MLPA. *Eur J Hum Genet* 17, 1439–1447.

Anna Bratuś. (2012) *DM Domain Genes: Sexual and Somatic Development During Vertebrate Embryogenesis, Embryogenesis*. Dr. Ken-Ichi Sato (Ed.), ISBN: 978-953-51-0466-7, InTech, Available from: http://www.intechopen.com/books/embryogenesis/dm-domain-genes-sexual-and-somatic-development-duringvertebrate-embryogenesis.

Cartegni L, Wang J, Zhu Z, Zhang MQ & Krainer AR. (2003) ESEfinder: a web resource to identify exonic splicing enhancers. *Nucleic Acids Res* 31, 3568–3571.

Cheng HH, Ying M, Tian YH, Guo Y, McElreavey K & Zhou RJ. (2006) Transcriptional diversity of DMRT1 (dsx- and mab-3-related transcription factor 1) in human testis. *Cell Res* 16, 389–393.

Davis RL, Homer VM, George PM & Brennan SO. (2009) A deep intronic mutation in FGB creates a consensus exonic splicing enhancer motif that results in afibrinogenemia caused by aberrant mRNA splicing, which can be corrected in vitro with antisense oligonucleotide treatment. *Hum Mutat* 30, 221–227.

Desmet FO, Hamroun D, Lalande M, Collod-Beroud G, Claustres M & Beroud C. (2009) Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res* 37, e67.

Erdman SE, Chen HJ & Burtis KC. (1996) Functional and genetic characterization of the oligomerization and DNA binding properties of the Drosophila doublesex proteins. *Genetics* 144, 1639–1652.

Genomes Project Consortium, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT & McVean GA. (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature* 491, 56–65.

Goymer P. (2007) Synonymous mutations break their silence. *Nat Rev Genet* 8, 92.

Jungwirth A, Diemer T, Dohle G, Giwercman A, Kopa Z, Krausz C & Tournaye H. (2015) *Guidelines on Male Infertility*. European Association of Urology (EAU), http://uroweb.org/guideline/male-infertility/.

Kallio M, Chang Y, Manuel M, Alastalo TP, Rallu M, Gitton Y, Pirkkala L, Loones MT, Paslaru L, Larney S, Hiard S, Morange M, Sistonen L & Mezger V. (2002) Brain abnormalities, defective meiotic chromosome synapsis and female subfertility in HSF2 null mice. *EMBO J* 21, 2591–2601.

Koressaar T & Remm M. (2007) Enhancements and modifications of primer design program Primer3. *Bioinformatics* 23, 1289–1291.

Ledig S, Hiort O, Scherer G, Hoffmann M, Wolff G, Morlot S, Kuechler A & Wieacker P. (2010) Array-CGH analysis in patients with syndromic and non-syndromic XY gonadal dysgenesis: evaluation of array CGH as diagnostic tool and search for new candidate loci. *Hum Reprod* 25, 2637–2646.

Ledig S, Hiort O, Wunsch L & Wieacker P. (2012) Partial deletion of DMRT1 causes 46, XY ovotesticular disorder of sexual development. *Eur J Endocrinol* 167, 119–124.

Lei N & Heckert LL. (2002) Sp1 and Egr1 regulate transcription of the Dmrt1 gene in Sertoli cells. *Biol Reprod* 66, 675–684.

Lima AC & Lopes AM. (2014) Autosomal mutations and spermatogenic failure. In: eLS. John Wiley & Sons, Ltd: Chichester. DOI: 10.1002/9780470015902.a0025310.

Liu HX, Zhang M & Krainer AR. (1998) Identification of functional exonic splicing enhancer motifs recognized by individual SR proteins. *Genes Dev* 12, 1998–2012.

Lopes AM, Aston KI, Thompson E, Carvalho F, Goncalves J, Huang N, Matthiesen R, Noordam MJ, Quintela I, Ramu A, Seabra C, Wilfert AB, Dai JC, Downie JM, Fernandes S, Guo XJ, Sha JH, Amorim A, Barros A,

Carracedo A, Hu ZB, Hurles ME, Moskovtsev S, Ober C, Paduch DA, Schiffman JD, Schlegel PN, Sousa M, Carrell DT & Conrad DF (2013) Human spermatogenic failure purges deleterious mutation load from the autosomes and both sex chromosomes, including the gene DMRT1. *PLoS Genet* 9(3), e1003349. doi:10.1371/journal.pgen.1003349.

Machado AZ, da Silva TE, Costa EMF, dos Santos MG, Nishi MY, Brito VN, Mendonca BB & Domenice S. (2012) Absence of inactivating mutations and deletions in the DMRT1 and FGF9 genes in a large cohort of 46, XY patients with gonadal dysgenesis. *Eur J Med Genet* 55, 690–694.

Matson CK, Murphy MW, Sarver AL, Griswold MD, Bardwell VJ & Zarkower D. (2011) DMRT1 prevents female reprogramming in the postnatal mammalian testis. *Nature* 476, 101–104.

Moniot B, Berta P, Scherer G, Sudbeck P & Poulat F. (2000) Male specific expression suggests role of DMRT1 in human sex determination. *Mech Dev* 91, 323–325.

Mou LS, Wang YD, Li HG, Huang Y, Jiang T, Huang WR, Li ZS, Chen J, Xie J, Liu YC, Jiang ZM, Li XX, Ye JX, Cai ZM & Gui YT. (2013) A dominant-negative mutation of HSF2 associated with idiopathic azoospermia. *Hum Genet* 132, 159–165.

Murphy MW, Sarver AL, Rice D, Hatzi K, Ye K, Melnick A, Heckert LL, Zarkower D & Bardwell VJ. (2010) Genome-wide analysis of DNA binding and transcriptional regulation by the mammalian Doublesex homolog DMRT1 in the juvenile testis. *Proc Natl Acad Sci U S A* 107, 13360–13365.

Õunap K, Uibo O, Zordania R, Kiho L, Ilus T, Oiglane-Shlik E & Bartsch O. (2004) Three patients with 9p deletions including DMRT1 and DMRT2: a girl with XY complement, bilateral ovotestes, and extreme growth retardation, and two XX females with normal pubertal development. *Am J Med Genet A* 130A, 415–423.

Pereira R, Phillips C, Pinto N, Santos C, dos Santos SE, Amorim A, Carracedo A & Gusmao L. (2012) Straightforward inference of ancestry and admixture proportions through ancestry-informative insertion deletion multiplexing. *PLoS ONE* 7, e29684.

Practice Committee of American Society for Reproductive Medicine in Collaboration with Society for Male Reproduction and Urology. (2008) Evaluation of the azoospermic male. *Fertil Steril* 90, S74–S77.

R Development Core Team. (2012) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Raymond CS, Parker ED, Kettlewell JR, Brown LG, Page DC, Kusz K, Jaruzelska J, Reinberg Y, Flejter WL, Bardwell VJ, Hirsch B & Zarkower D. (1999) A region of human chromosome 9p required for testis development contains two genes related to known sexual regulators. *Hum Mol Genet* 8, 989–996.

Smale ST & Baltimore D. (1989) The initiator as a transcription control element. *Cell* 57, 103–113.

Tannour-Louet M, Han S, Corbett ST, Louet JF, Yatsenko S, Meyers L, Shaw CA, Kang SHL, Cheung SW & Lamb DJ. (2010) Identification of de novo copy number variants associated with human disorders of sexual development. *PLoS ONE* 5, e15392.

Tewes AC, Ledig S, Tuttelmann F, Kliesch S & Wieacker P. (2014) DMRT1 mutations are rarely associated with male infertility. *Fertil Steril* 102, 271–816820 e3.

Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M & Rozen SG. (2012) Primer3–new capabilities and interfaces. *Nucleic Acids Res* 40, e115.

Veitia R, Nunes M, Brauner R, DocoFenzy M, JoannyFlinois O, Jaubert F, LortatJacob S, Fellous M & McElreavey K. (1997) Deletions of distal 9p associated with 46, XY male to female sex reversal: definition of the breakpoints at 9p23.3–p24.1. *Genomics* 41, 271–274.

Zarkower D. (2013) DMRT genes in vertebrate gametogenesis. *Curr Top Dev Biol* 102, 327–356.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Figure S1. Principal component analysis of reference populations and NOA patient tested using AIMs.
Figure S2. SNaPshot profiles of the heterozygous state for all variants included in the multiplex reaction.
Figure S3. Snapshot of the alignment of mammalian and avian DMRT1 sequences.
Table S1. Primers designed for *DMRT1* amplification and sequencing.
Table S2. Summary of novel and rare variants tested in Portuguese patients and controls.
Table S3. SNaPshot multiplex design.
Table S4. DMRT1 variants retrieved from Ensembl database.
Table S5. TFBSs predicted for reference or altered human sequences in the conserved motif in the DMRT1 promoter sequence.
Table S6. TFBS motifs overlaping with variants in the 5′-UTR.
Table S7. DMRT1 cDNA variants detected in large genome sequencing projects retrived from Ensembl database.
Data S1. Extended materials and methods.

Paper III - The mutational spectrum of *WT1* in male infertility

The Journal of Urology, 2015

The Mutational Spectrum of *WT1* in Male Infertility

Catarina M. Seabra, Sofia Quental, Ana C. Lima, Filipa Carvalho, João Gonçalves, Susana Fernandes, Iris Pereira, Júlia Silva, Patrícia I. Marques, Mário Sousa, Alberto Barros, Susana Seixas, António Amorim and Alexandra M. Lopes*

From the Institute of Molecular Pathology and Immunology (CMS, SQ, ACL, PIM, SS, AA, AML), Graduate Program in Areas of Basic and Applied Biology (ACL), Department of Genetics, Faculty of Medicine (FC, SF, AB), Laboratory of Cell Biology, Unit for Multidisciplinary Research in Biomedicine (MS), Abel Salazar Institute of Biomedical Sciences (ACL, PIM) and Faculty of Sciences (AA), University of Porto, Porto, Health Sciences Autonomous Section, University of Aveiro (CMS), Aveiro and Department of Human Genetics, National Institute of Health Dr. Ricardo Jorge (JG, IP, JS), Lisboa, Portugal, and Department of Genetics, Washington University School of Medicine (ACL), St. Louis, Missouri

Purpose: We evaluated the impact of *WT1* mutations in isolated severe spermatogenic impairment in a population of European ancestry. *WT1* was first identified as the gene responsible for Wilms tumor. It was later associated with a plethora of clinical phenotypes often accompanied by urogenital defects and male infertility. The recent finding of *WT1* missense mutations in Chinese azoospermic males without major gonadal malformations broadened the phenotypic spectrum of *WT1* defects and motivated this study.

Materials and Methods: We analyzed the *WT1* coding region in a cohort of 194 Portuguese patients with nonobstructive azoospermia and in 188 with severe oligozoospermia with increased depth for the exons encoding the regulatory region of the protein. We also analyzed a group of 31 infertile males with a clinical history of unilateral or bilateral cryptorchidism and 1 patient with anorchia.

Results: We found 2 *WT1* missense substitutions at higher frequency in patients than in controls. 1) A novel variant in exon 1 (p.Pro130Leu) that disrupted a mammalian specific polyproline stretch in the self-association domain was more frequent in azoospermia cases (0.27% vs 0.13%, $p = 0.549$). 2) A rare variant in a conserved residue in close proximity to the first zinc finger (pCys350Arg) was more frequent in severe oligozoospermia cases (0.80% vs 0.13%, $p = 0.113$).

Conclusions: Results suggest a role for rare *WT1* damaging variants in severe spermatogenic failure in populations of European ancestry. Large multicenter studies are needed to fully assess the contribution of *WT1* genetic alterations to male infertility in the absence of other disease phenotypes.

Key Words: testis; infertility, male; genes, Wilms tumor; mutation; European continental ancestry group

Abbreviations and Acronyms

DDS = Denys-Drash

NOA = nonobstructive azoospermia

NR5A1 = nuclear receptor steroidogenic factor 1

WT1 = Wilms tumor 1

Accepted for publication November 5, 2014.

Study received National Institute of Health Dr. Ricardo Jorge ethics committee approval.

Supported by the Portuguese Foundation for Science and Technology FCT/MCTES (PIDDAC) and European funds (FEDER) through the COMPETE program, Research Grant PTDC/SAU-GMG/101229/2008 (AML), CIGMH, Pest-OE/SAU/UI00009/2011, FCT, and FCT Fellowships SFRH/BPD/73366/2010 (AML), SFRH/BPD/64025/2009 (SQ), SFRH/BD/51695/2011 (ACL) and SFRH/BD/68940/2010 (PIM).

* Correspondence: Institute of Molecular Pathology and Immunology, University of Porto, Porto, Portugal (telephone: 00351225570700; e-mail: alopes@ipatimup.pt).

THE *WT1* gene located at 11p13 encodes a protein with a C-terminal zinc finger domain that is involved in DNA and RNA binding. WT1 acts as a transcription factor through the interaction of WT1 activation and repression domains with its targets.[1,2] This activity is thought to be modified by dimerization of WT1 with other proteins (heterodimerization) or with itself (homodimerization) at the N-terminal self-association domain.

0022-5347/15/1935-1709/0
THE JOURNAL OF UROLOGY®

http://dx.doi.org/10.1016/j.juro.2014.11.004
Vol. 193, 1709-1715, May 2015
Printed in U.S.A.

Genetic defects in *WT1* typically result in 1 of 3 congenital syndromes, including WAGR (Wilms tumor, aniridia, genitourinary alteration and mental retardation), DDS or Frasier syndrome, characterized by malformation of the gonadal ridge (gonadal dysgenesis, hypospadias and cryptorchidism) and kidneys (horseshoe kidney and renal hypoplasia).[3] This association is in accord with the preponderant role of *WT1* in urogenital system development and differentiation.

WT1 is a crucial factor in male sex determination with an essential role in the male gonadal differentiation pathway. This is supported by the fact that gonads fail to develop in $Wt1^{-/-}$ mice.[4] Moreover, several lines of evidence demonstrate its importance in different stages of testicular and germ cell development. Chang et al reported that WT1 dependent suppression of WNT/β-catenin signaling in Sertoli cells is essential for the normal development of primordial germ cells.[5] Studies in which *Wt1* was specifically inactivated in Sertoli cells at different developmental stages revealed its importance in testicular differentiation[6] and in the maintenance of Sertoli cell polarity and spermatogenesis in adulthood.[7] Interestingly the phenotype of germ cell loss observed in *Wt1* conditional knockout mice resembles that observed in Sertoli-cell only syndrome and in accord several missense mutations in the *WT1* gene were recently described in a cohort of Chinese patients with NOA.[7]

Phenotypic expression of *WT1* defects varies and depends on the affected protein domains. Patients with C-terminal missense or nonsense mutations typically show severe gonadal dysgenesis and/or nephropathy resulting from a dominant negative action of heterozygous *WT1* missense mutations or from haploinsufficiency.[3,8,9] Most mutations reported in the first exons of this gene result in truncated proteins and were found in patients with renal tumor and genitourinary abnormalities. In fact, missense mutations affecting only the WT1 protein N-terminus are expected to have a milder impact on its physiological function and result in milder gonadal malformations since the DNA binding domain should remain intact.[10] Accordingly most mutations found to date in infertile patients without major disturbances of testicular development are located in the WT1 protein N-terminus.[7]

We resequenced the *WT1* coding region in a cohort of Portuguese patients, focusing on the first 6 exons, which encode the regulatory domain. To determine the impact of the 2 identified missense *WT1* variants we 1) reassessed the protein conservation across vertebrates and 2) reviewed the spectrum of *WT1* mutations affecting male fertility and compared them to those in control populations from large genome sequencing projects.

MATERIALS AND METHODS

Patients and Controls

DNA samples from peripheral blood leukocytes of 194 men with NOA and 188 with severe oligozoospermia (fewer than 1 million sperm per ml) who had idiopathic spermatogenic failure were collected at the Human Genetics Department, National Institute of Health Dr. Ricardo Jorge, and the Genetics Department, Faculty of Medicine, University of Porto, where routine molecular diagnosis was done for male infertility. A group of 31 infertile males with a clinical history of unilateral or bilateral cryptorchidism and 1 patient with anorchia were also selected after physical examination, hormonal testing (follicle-stimulating hormone and testosterone) and standard clinical genetic screening for karyotypic anomalies and Y chromosome microdeletions. Patients with a known cause of infertility were excluded from study. Molecular studies were performed in coded DNA samples after receiving informed consent.

As controls we obtained DNA from peripheral blood from 373 Portuguese men, including 72 with normozoospermia (normal sperm parameters) and 301 who had fathered at least 1 child. This study was included in the project, Copy Number Variation in Infertile Men Genomic Regions: Screening in the Portuguese Population (PTDC/SAU-GMG/101229/2008), which was approved by the National Institute of Health Dr. Ricardo Jorge ethics committee.

Analysis

WT1 coding sequence. Supplementary table 1 (http://jurology.com/) lists the *WT1* primers used in this study. To analyze the *NR5A1* sequence we used the primers described by Bashamboo et al.[11] DNA fragments were amplified and sequenced, and all putative variants were individually confirmed.

Restriction fragment length polymorphism. This technique was applied to screen Portuguese controls for the c.1048T>C variant in *WT1* exon 6. The restriction endonuclease *Cfr42I* was used to cleave a 1,020 bp amplicon into 2 fragments of 832 and 188 bp, respectively, which were resolved by electrophoresis in polyacrylamide gel.

In silico. WT1 protein sequences of several species of mammals and from chickens were retrieved from the Ensembl database (http://www.ensemble.org/) and manually curated and aligned with the ClustalW algorithm in Geneious v.5.5.8 (http://www.geneious.com/). PolyPhen-2 (Polymorphism Phenotyping, version 2)[12] (http://genetics.bwh.harvard.edu/pph2/) and SIFT (http://sift.jcvi.org/) were used to predict the impact of nonsynonymous substitutions in *WT1* patients and controls. As a reference we used the Ensembl WT1 protein ENSP00000331327 (transcript sequence ENST00000332351). This transcript comprises all 10 *WT1* exons (3,122 bp) and initiates with the upstream CUG codon. We retrieved variant data from 1000 Genomes (http://www.1000-genomes.org/),[13] NHLBI (National Heart, Lung and Blood Institute) GO (Grand Opportunity) ESP (Exome Sequencing Project)

(http://evs.gs.washington.edu/EVS/) and CLINSEQ® (supplementary table 2, http://jurology.com/).

RESULTS

Mutation Screening

Mutation screening was performed in 2 stages. Initially the whole *WT1* coding region (fig. 1) and flanking intronic regions were sequenced in 92 patients with NOA. One new missense variant was found in the first exon of the gene c.389C>T (p.Pro130Leu, ENST00000332351). This finding was in agreement with our hypothesis that this cohort would be enriched for rare damaging variants in the N-terminal region of the protein. Therefore, at stage 2 we increased the coverage of the first exons (1 to 6) of the *WT1* gene, which were sequenced in a larger number of patients. Overall this region was analyzed in 169 patients with NOA and 2 missense variants were found, including 1 in exon 1, resulting in a proline to leucine substitution (p.Pro130Leu, c.389C>T, ENST00000332351), and 1 variant in exon 6, resulting in a cysteine to arginine substitution (p.Cys350Arg, c.1048T>C, ENST00000332351). The patient carrying the p.Cys350Arg variant had been referred to the fertility clinic with a primary diagnosis of NOA. However, spermatozoa were recovered from the ejaculate on a second attempt and a final diagnosis of severe oligozoospermia was established. We also screened 31 patients with unilateral or bilateral cryptorchidism and 1 with anorchia for *WT1* coding mutations and detected the rare p.Cys350Arg in the patient with anorchia but found no additional alterations. Patients harboring *WT1* coding variants, including 1 with azoospermia and 1 with severe oligozoospermia as well as the patient with anorchia, were also screened for exonic as well as proximal flanking intronic mutations in *NR5A1*, which is associated with severe spermatogenic impairment.[11,14] No alteration was detected.

Overall *WT1* exon 1 was sequenced in 183 patients and 373 Portuguese controls, of whom all had azoospermia, including 301 who were fertile and 72 with normozoospermia. The p.Pro130Leu variant was found in heterozygosity in 1 patient and in 1 fertile control with an unknown sperm count. Exon 6 was sequenced in 194 patients with severe spermatogenic impairment, including 193 with NOA and 1 with severe oligozoospermic. The p.Cys350Arg variant was then tested by restriction fragment length polymorphism in 188 patients with severe oligozoospermia and 371 controls, including 299 who were fertile and 72 with normozoospermia. It was detected in heterozygosity in 3 patients and in 1 fertile man with an unknown sperm count. The allele frequency of each missense variant was higher in patients than in controls. The p.Pro130Leu variant was twice more frequent in patients with azoospermia than in controls (0.27% vs 0.13%). The p.Cys350Arg variant showed a greater than sixfold difference in frequency between patients with severe oligozoospermia and controls (0.80% vs 0.13%), although these differences did not attain statistical significance (Fisher exact test, $p = 0.549$ and 0.113, respectively). None of these variants was present in individuals with a known sperm count. The p.Cys350Arg substitution in exon 6 (rs142059681) was previously detected in the large-scale ESP at low frequency in European-American and African-American men (0.08% and 0.02%, respectively).

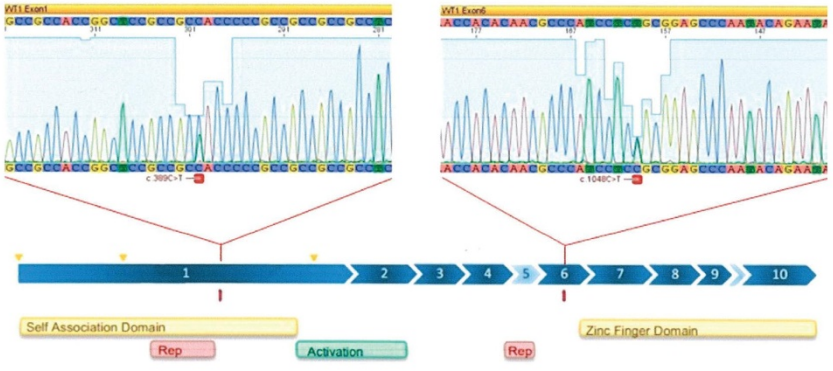


Figure 1. *WT1* gene exons and corresponding functional domains of encoded protein. Electropherogram shows c.389C>T and c.1048C>T substitutions (underlined) in exons 1 and 6, respectively. Exon 5 is alternatively spliced as well as KTS insertion between exons 9 and 10 (light blue). Yellow triangles represent 3 alternative transcription start sites, of which middle site is canonic.

Coding Variant Functional Impact

Two missense variants were detected at higher frequency in our group of patients with severe spermatogenic impairment than in controls in the same population. In 3 men with severe oligozoospermia and 1 with anorchia we detected a rare variant in exon 6 (rs142059681). This variant resulted in a substitution predicted to be damaging by PolyPhen-2 (score 0.999) and SIFT (score 0), in which a conserved cysteine residue is replaced by an arginine (p.Cys350Arg). This substitution in close proximity to the first zinc finger may interfere with the stabilization of this important functional domain.

The second variant, which to our knowledge we report for the first time, was located in exon 1 and altered the encoded amino acid from proline to leucine (p.Pro130Leu c.389C>T, ENST00000332351). This disrupted a polyproline stretch in the WT1 self-association domain. In silico prediction of the functional impact of this substitution using PolyPhen-2 as well as the fact that the protein isoform started at the major initiation site classified this variant as possibly damaging (score 0.770). Notably in the absence of a tridimensional structure of this region of the protein PolyPhen-2 only considers phylogenetic conservation of this residue and the mutations annotated in this position known to cause Mendelian disease. The SIFT score also supported a deleterious impact on the protein, although with low confidence (score 0).

To better evaluate the functional impact of the variants in our patients we analyzed WT1 conservation by retrieving and aligning all high quality sequences from available mammalian species, and an avian and a fish sequence from Ensembl, release 75, February 2014 (fig. 2). Although the C-terminus of WT1 (harboring the 4 Kruppel-like zinc fingers) is more highly conserved across species than the transregulatory domain, as previously noted,[15,16] the conservation of other regions in the N-terminus of the protein was also remarkable. Indeed, most of the activation domain, a block of 19 amino acids in the self-association domain and the C-terminal part of exon 6 juxtaposed to the zinc finger domain, are completely conserved in mammals and also in amniotes. Some mammalian specific features also stand out in the alignment, such as a polyglycine and a polyproline stretch in the N-terminus as well as the alternatively spliced exon 5.[17] The polyproline stretch is highly conserved in placental and marsupial mammals with a minimum of 6 consecutive proline residues in all species of this group.

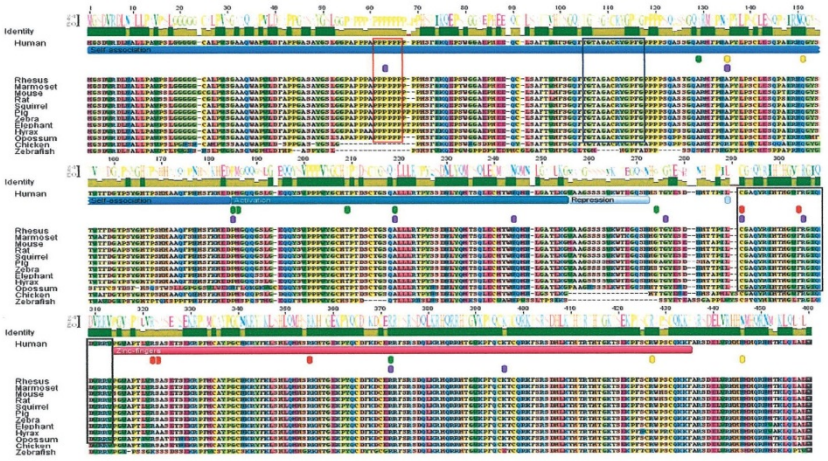


Figure 2. Sequence downstream of main initiator AUG site of WT1 protein conservation and alignment of human WT1 sequences with mammalian orthologues, chickens and zebrafish. High level of amino acid similarity was evident at C-terminus domain across vertebrates. In N-terminal region most of activation domain, block of 19 amino acids in self-association domain (blue outline) and C-terminal part of exon 6 juxtaposed to zinc finger domain (black outline) are completely conserved in mammals and amniotes. Some mammalian specific features that stand out in alignment are polyglycine and polyproline stretch in N-terminus (red outline) and alternatively spliced exon 5 at positions 265 to 272. Below human sequence note previously reported mutations implicated in nonsyndromic male infertility (purple areas) and missense variants in controls from large sequencing projects in Ensembl. PolyPhen-2 predictions included probably damaging (red symbols), possibly damaging (yellow symbols) and benign (green symbols) (supplementary table 2, http://jurology.com/).

WT1 Variant Pathological Mutations and Coding Polymorphisms, and Male Fertility

To grasp the potential impact of coding variants in the protein we compared the spectrum of missense *WT1* variants in controls in large sequencing projects to that of substitutions in patients with impaired fertility (see table).[3,11,18,19]

Apart from 3 well-defined syndromes (see Appendix), *WT1* mutations were also identified in 46,XY male patients with genital abnormalities (typically gonadal dysgenesis, hypospadias and/or cryptorchidism) in whom Wilms tumor may develop but who do not present with the nephropathy typical of DDS or Frasier syndrome.[3,10,18–21] Mutations in these patients cluster in the initial exons of the gene and are missense or nonsense. In the latter case they lead to prematurely truncated proteins. Two missense mutations were previously reported in patients with isolated genital and gonadal malformations such as hypospadias and cryptorchidism, and were considered pathogenic,[10,18,21] including 1 in the self-association domain (p.Ala199Thr) and another in the activation domain (p.Pro249Ser). Surprisingly these variants were present in unphenotyped controls of European ancestry with p.Ala199Thr attaining frequency that was not so negligible for a deleterious variant (rs9332973, 1.3% in CEU).

Functional predictions for this variant differed depending on the algorithm used. According to PolyPhen-2 it is benign (0.172) but the SIFT score is within the range expected for a deleterious substitution (0.02). The p.Pro249Ser substitution, which was also found in a patient with XY ambiguous genitalia,[19] was rare in controls at only 0.06% in European-American men in ESP (rs2234584). It was classified as nondamaging by PolyPhen-2 and SIFT (0.041 and 0.16, respectively). Although we could not exclude that some of these controls may have had genital or gonadal malformations, the relatively high frequency of p.Ala199Thr allied to the bioinformatics predictions suggests that it is more likely a risk factor for the observed phenotype rather than a highly penetrant pathogenic variant.

The missense variants recently associated with NOA in Chinese patients are equally distributed throughout the activation domain, exon 6 and the zinc finger region. The 2 substitutions in the zinc fingers (p.Arg430Gln and p.Lys454Arg) are predicted to be damaging (PolyPhen-2 score 0.994 and 0.807, respectively). They affect amino acids conserved in amniotes and in vitro they interfered with the function of the protein.[7] All 5 missense variants in the zinc fingers in controls were rare (minor allele frequency less than 0.1%) and only 1 was benign, reflecting the strong functional constraints in this region of the protein. In exon 6 the p.Arg363Ser substitution found by Wang et al in 1 man with azoospermia[7] and p.Cys350Arg, which is more common in Portuguese patients with oligozoospermia than in fertile controls, affect positions that are highly conserved in vertebrates and classified by PolyPhen-2 as probably damaging. The p.Gly338Ala substitution also described in azoospermia is conserved across amniotes and predicted to be benign. Only 2 rare missense variants (0.01% minor allele frequency) were found in this exon in controls apart from p.Cys350Arg and their functional impact differs. Ser336Asn (rs371021920) in the first amino acid of exon 6 is predicted not to affect protein function and it is a fairly variable position even in mammals. In contrast, Phe362Cys

WT1 genetic defects detected in male patients with clinical presentation, including isolated nonsyndromic genital anomalies, morphological gonadal defects or spermatogenic failure

References	Amino Acid Change	Exon	Phenotype	Affected Functional Domain	Ensemble *WT1* Transcript ENST00000332351 PolyPhen2	SIFT
N-terminus:						
Köhler et al[18]	Val130Term	E1	Hypospadias, unilat cryptorchidism	Self-association domain	Not applicable	
Wang et al[21]	Ala199Thr	E1	Hypospadias	Self-association domain	0.172	0.02
Köhler et al[18,*]	Pro249Ser	E2	Hypospadias, unilat cryptorchidism		0.041	0.16
Wang et al[7]	Ala282Pro, Asn307Ser, Gly338Ala, Arg363Ser	E3, E4, E6, E6	Nonobstructive azoospermia	Activation domain, activation domain, —, —	0.983, 0.994, 0.075, 0.996	0, 0, 0.06, 0.01
C-terminus:						
Wang et al[7]	Arg430Gln, Lys454Arg	E8, E9	Nonobstructive azoospermia	Zinc fingers (nonDNA contacting residue)	0.994, 0.807	0.16, 0.12
Köhler et al: Pediatr Res 1999; **45:** 187	Arg430Term	E8	Hypospadias, bilat cryptorchidism, Wilms tumor	Zinc fingers	Not applicable	
Köhler et al[19]	Arg458Term	E9	Severe hypospadias, bilat cryptorchidism, glomerulosclerosis	Zinc fingers	Not applicable	

* Also detected in patient with XY ambiguous genitalia and congenital heart disease.[10]

(rs150194429) in the C-terminal end of exon 6 is predicted to be damaging. The different conservation level and predicted functional impact of variants in the N-terminal and C-terminal of exon 6 strongly suggest that the latter may be crucial for protein function, likely due to its proximity to the first zinc finger.

DISCUSSION

Most *WT1* substitutions reported to date are associated with severe phenotypes such as cryptorchidism, hypospadias, ambiguous genitalia, syndromic complications and renal tumor,[3,9,22,23] and are located in the zinc finger region of the *WT1* gene. More recently *WT1* missense substitutions were reported in Chinese infertile males without syndromic manifestations or other major gonadal and/or urogenital abnormalities.[7] Four of 6 of the latter variants are clustered in the N-terminal region of the protein, suggesting that missense substitutions in this region of the protein result in milder impairment of gonadal function.

We screened the *WT1* coding region and the flanking intronic sequence for mutations in a cohort of Portuguese males with NOA who were previously analyzed for genome-wide copy number variation[24] and in men with oligozoospermia. We focused on the N-terminal region of the gene and included the first exon, which was not efficiently captured in a previous mutation screen due to its high GC content.[7]

We identified 2 missense variants at higher frequency in our infertile patients than in geographically matched controls. These variants were absent in men with a known sperm count. In exon 1 p.Pro130Leu, which to our knowledge we report for the first time, disrupts a polyproline stretch in the self-association domain, which is involved in transrepression of WT1 target genes.[25] All previously reported mutations in this exon lead to truncated and, thus, dysfunctional WT1 proteins associated with severe phenotypes.[8,18] This substitution was found in 1 patient with azoospermia as well as in 1 Portuguese fertile control with an unknown sperm count. This variant appears to be population specific and its pathogenic potential should be evaluated by analyzing a larger cohort of European men with azoospermia and matched controls with normozoospermia.

The second substitution (p.Cys350Arg, rs142059681) was identified in patients with severe oligozoospermia with a greater than sixfold difference in frequency compared to controls. This variant was also present in a patient with anorchia who harbored no *NR5A1* coding mutation. p.Cys350Arg was also detected in controls of European and African ancestry. Therefore, it is a recurrent mutation or a relatively old, mildly deleterious allele that likely represents a risk factor for spermatogenic failure.

By surveying available data on *WT1* missense variants in control populations we concluded that the rare variants with a higher probability of damaging the protein are located in the C-terminus of exon 6 and in the zinc fingers, in agreement with the strong functional constraints in these regions. The analysis of WT1 orthologues confirmed high conservation of the C-terminal region of exon 6 in vertebrates. The close proximity to the first zinc finger suggests that this region may have a role in stabilizing this crucial DNA binding element.

The *WT1* variants associated with azoospermia in Chinese patients[7] as well as those that we report are clustered in the N-terminus of the protein. Another 2 substitutions in the *WT1* N-terminus were previously reported as pathogenic, including 1 in the self-association domain (p.Ala199Thr) and 1 in the activation domain (p.Pro249Ser). They were also found in controls with the former substitution attaining a frequency of 1.3%, which is not so negligible. The pathogenic potential of these variants must be further evaluated, considering that rather than having a strong effect on heterozygosity they may instead confer a risk of impaired spermatogenesis.

CONCLUSIONS

Available evidence indicates that *WT1* N-terminal missense variants result in milder impairment of gonadal and kidney development, as manifested by isolated spermatogenic failure in otherwise healthy men, or in increased risk of gonadal or genital malformations, such as hypospadias or cryptorchidism. These results call for large multicenter studies to fully assess the contribution of *WT1* genetic alterations to male infertility, establish clear genotype-phenotype correlations and determine the usefulness of *WT1* mutation screening in patients of European ancestry.

ACKNOWLEDGMENTS

Drs. Ana Aguiar, Joaquim Nunes, Ana Paula Soares and Carlos Calhaz-Jorge, Unidade de Medicina da Reprodução, Departamento de Obstetrícia, Ginecologia e Medicina da Reprodução, CHLN-Hospital de Santa Maria, Lisboa, and Graça Pinto and Sónia Correia, Unidade de Medicina da Reprodução, Maternidade, assisted with the study. Dr. Alfredo da Costa performed patient clinical evaluation and sperm analysis.

APPENDIX

WT1 associated syndromes

	WAGR	Denys-Drash	Frasier
Genotype	Large 11p13 deletions: *WT1* gene: genitourinary features *PAX6* gene: aniridia Tumors usually carry intragenic somatic mutations in remaining *WT1* allele.	Heterozygous point mutations in exons coding ZF region of *WT1*.	Constitutional intronic mutations in second donor splice site of intron 9 on 1 *WT1* copy.
Functional Consequences	Individuals with constitutional *WT1* deletion present at high risk for Wilms tumor (greater than 20%) and should be monitored.	DDS mutant *WT1* proteins fail to bind DNA and act in dominant negative fashion by forming homodimers with normal WT1 protein. Prevents WT1 physiological activity.	Prevention of production of KTS containing isoforms. There is shift in KTS isoform ratio leading to imbalance of WT1 isoform functions rather than formation of mutant protein.
Clinical phenotype	Wilms tumor Aniridia Genitourinary anomalies Mental retardation	DDS triad: Mesangial sclerosis (nephropathy) Genital abnormalities (mild to XY pseudo-hermaphroditism) Wilms tumor	XY pseudohermaphroditism Glomerulonephropathy Renal tumor usually does not develop

REFERENCES

1. Discenza MT and Pelletier J: Insights into the physiological role of WT1 from studies of genetically modified mice. Physiol Genomics 2004; **16:** 287.
2. Scharnhorst V, van der Eb AJ and Jochemsen AG: WT1 proteins: functions in growth and differentiation. Gene 2001; **273:** 141.
3. Little M and Wells C: A clinical overview of WT1 gene mutations. Hum Mutat 1997; **9:** 209.
4. Kreidberg JA, Sariola H, Loring JM et al: WT-1 is required for early kidney development. Cell 1993; **74:** 679.
5. Chang H, Gao F, Guillou F et al: Wt1 negatively regulates beta-catenin signaling during testis development. Development 2008; **135:** 1875.
6. Gao F, Maiti S, Alam N et al: The Wilms tumor gene, Wt1, is required for Sox9 expression and maintenance of tubular architecture in the developing testis. Proc Natl Acad Sci U S A 2006; **103:** 11987.
7. Wang XN, Li ZS, Ren Y et al: The Wilms tumor gene, Wt1, is critical for mouse spermatogenesis via regulation of Sertoli cell polarity and is associated with non-obstructive azoospermia in humans. PLoS Genet 2013; **9:** e1003645.
8. Huff V: Wilms' tumours: about tumour suppressor genes, an oncogene and a chameleon gene. Nat Rev Cancer 2011; **11:** 111.
9. Royer-Pokora B, Beier M, Henzler M et al: Twenty-four new cases of WT1 germline mutations and review of the literature: genotype/phenotype correlations for Wilms tumor development. Am J Med Genet A 2004; **127A:** 249.
10. Köhler B, Pienkowski C, Audran F et al: An N-terminal WT1 mutation (P181S) in an XY patient with ambiguous genitalia, normal testosterone production, absence of kidney disease and associated heart defect: enlarging the phenotypic spectrum of WT1 defects. Eur J Endocrinol 2004; **150:** 825.
11. Bashamboo A, Ferraz-de-Souza B, Lourenço D et al: Human male infertility associated with mutations in NR5A1 encoding steroidogenic factor 1. Am J Hum Genet 2010; **87:** 505.
12. Adzhubei IA, Schmidt S, Peshkin L et al: A method and server for predicting damaging missense mutations. Nat Methods 2010; **7:** 248.
13. Altshuler DL, Durbin RM, Abecasis GR et al: A map of human genome variation from population-scale sequencing. Nature 2010; **467:** 1061.
14. Ropke A, Tewes AC, Gromoll J et al: Comprehensive sequence analysis of the NR5A1 gene encoding steroidogenic factor 1 in a large group of infertile males. Eur J Hum Genet 2013; **21:** 1012.
15. Kent J, Coriat AM, Sharpe PT et al: The evolution of WT1 sequence and expression pattern in the vertebrates. Oncogene 1995; **11:** 1781.
16. Miles C, Elgar G, Coles E et al: Complete sequencing of the Fugu WAGR region from WT1 to PAX6: dramatic compaction and conservation of synteny with human chromosome 11p13. Proc Natl Acad Sci U S A 1998; **95:** 13068.
17. Natoli TA, McDonald A, Alberta JA et al: A mammal-specific exon of WT1 is not required for development or fertility. Mol Cell Biol 2002; **22:** 4433.
18. Köhler B, Biebermann H, Friedsam V et al: Analysis of the Wilms' tumor suppressor gene (WT1) in patients 46,XY disorders of sex development. J Clin Endocrinol Metab 2011; **96:** E1131.
19. Köhler B, Schumacher V, l'Allemand D et al: Germline Wilms tumor suppressor gene (WT1) mutation leading to isolated genital malformation without Wilms tumor or nephropathy. J Pediatr 2001; **138:** 421.
20. Evaluation of the azoospermic male. Fertil Steril 2008; **90:** S74.
21. Wang Y, Li Q, Xu J et al: Mutation analysis of five candidate genes in Chinese patients with hypospadias. Eur J Hum Genet 2004; **12:** 706.
22. Baird PN, Santos A, Groves N et al: Constitutional mutations in the WT1 gene in patients with Denys-Drash syndrome. Hum Mol Genet 1992; **1:** 301.
23. Melo KFS, Martin RM, Costa EMF et al: An unusual phenotype of Frasier syndrome due to IVS9+4C > T mutation in the WT1 gene: predominantly male ambiguous genitalia and absence of gonadal dysgenesis. J Clin Endocrinol Metab 2002; **87:** 2500.
24. Lopes AM, Aston KI, Thompson E et al: Human spermatogenic failure purges deleterious mutation load from the autosomes and both sex chromosomes, including the gene DMRT1. PLoS Genet 2013; **9:** e1003349.
25. Madden SL, Cook DM and Rauscher FJ 3rd: A structure-function analysis of transcriptional repression mediated by the WT1, Wilms' tumor suppressor protein. Oncogene 1993; **8:** 1713.

Discussion

Non-obstructive azoospermia (NOA) is characterized by the absence of spermatozoa in the ejaculate, and is a type of SFF that affects ~1% of males worldwide (Practice Committee of American Society for Reproductive Medicine in collaboration with Society for Male Reproduction and Urology 2008). This number is not negligible, especially if we consider that this condition directly compromises the individual's reproductive fitness. Therefore, any highly penetrant dominant mutation with a severe impact on fertility should be rare and most likely attributed to *de novo* events. On the other hand, milder recessive variants may be maintained in the population for longer periods at low frequency. The phenotypic impact of the latter variants is more difficult to assess and a convincing association with disease may only be revealed in ethnic groups with high levels of inbreeding.

The genetics of SFF: a combination of rare mildly deleterious variants?

As previously mentioned, spermatogenesis is a highly coordinated process with many molecular players required to produce viable sperm. Unsurprisingly, thousands of genes control spermatogenesis progression and thus deleterious variants at any of these loci may be the underlying cause of spermatogenic impairment resulting in male infertility. Indeed, defects in a variety of genes have already been associated to severe phenotypes with gonadal malformation (Reviewed in Matzuk and Lamb (2008); Paper I). This is the case for the two genes studied in this chapter, *DMRT1* and *WT1*. Large deletions or heterozygous point mutations in patients with severe phenotypes may suggest a predominant model of haploinsufficiency for these genes in disorders of gonadal development (Huff 2011; Quinonez *et al.* 2013). However, recent data supports that genetic defects in these genes may be linked to less severe phenotypes of spermatogenic failure (Lopes *et al.* 2013; Wang *et al.* 2013; Tewes *et al.* 2014) suggesting different models with variable outcomes in terms of infertility phenotypes.

Hemizygous chromosomal deletions involving the *DMRT* cluster have been implicated in syndromic and non-syndromic forms of XY gonadal dysgenesis (GD) that show variability in both the extent of deleted *DMRT1* coding sequence and the phenotypic outcome. However, deletions in the vicinity but not encompassing *DMRT1* were reported (Barbaro *et al.* 2009; Tannour-Louet *et al.* 2010) in two cases of GD and no relevant variants were found in the *DMRT1* sequence in a cohort of 33 patients with GD (Machado *et al.* 2012). Recently, large *DMRT1* deletions were detected in NOA patients without gonadal abnormalities, thus

representing increased risk of SSF (Lopes *et al.* 2013). Also, considering that the normal allele is rarely sequenced for screening of smaller genetic alterations, hemizygosity of *DMRT1* might not be sufficient to cause GD. In this regard, the study presented in paper II supports a model where changes of *DMRT1* gene dosage through misregulation may contribute to NOA, reinforcing that rare variants affecting DMRT1 function may underlie some cases of SFF. This work describes novel and rare variants in Portuguese patients (Paper II - Tables 1 and 2) that are predicted to alter gene expression/function by disrupting transcription factor binding-sites (TFBS) for proteins highly and/or differentially expressed in testis or by altering mRNA splicing. *Per se*, these variants are likely to have a mild impact on spermatogenesis outcome and suggest that a combinatorial effect of cis-regulatory variants could be implicated in SFF. Moreover, the lack of coding mutations predicted to be damaging to the protein and the overall low protein sequence diversity, with highly conserved DM and DMRT1-like domains among vertebrates, indicates that this locus might be under strong functional constrains. This supports that variants with strong functional impact on spermatogenesis are not maintained in the population, reinforcing previous observations that *DMRT1* genetic defects are rare (Lopes *et al.* 2013; Tewes *et al.* 2014). Taken together, these data may suggest that the burden of deleterious genetic variants at the *DMRT1* locus is phenotypically expressed in two different modes. Complete loss of DMRT1 may result in more severe phenotypes presenting variable degrees of gonadal malformation, whereas the presence of one functional copy ensures normal gonad development but is likely insufficient for proper germ cell maintenance and differentiation. This would translate into a recessive model for *DMRT1* in cases of GD and one of haploinsufficiency in SFF. Although many mechanisms of disease expression differ in mice and humans, this hypothesis would fit the recessive model described in *DMRT1* KO mice (Raymond *et al.* 2000).

Most *WT1* alterations reported to date are associated with severe phenotypes of gonadal malformation (e.g. cryptorchidism, hypospadias and ambiguous genitalia) and renal tumor (Little and Wells 1997; Royer-Pokora *et al.* 2004). While some are attributed to large 11p13 deletions (Xu *et al.* 2008; Lopes *et al.* 2013), many phenotypes arise from point mutations in the *WT1* sequence. Interestingly, it appears that the phenotypic expression of these defects varies according to the affected protein domain (Table paper III). Heterozygous C-terminal missense or nonsense mutations in the ZF region of the *WT1* gene have either a dominant negative effect or lead to haploinsufficiency, and are typically linked to severe gonadal dysgenesis and/or nephropathy (Little and Wells 1997; Royer-Pokora *et al.* 2004; Huff 2011). On the other hand, missense substitutions associated with NOA mostly cluster in the N-

terminal region of the protein and have a milder effect on gonadal function (Wang *et al.* 2013). Indeed, the work reported in paper III is in agreement with these observations. Full genetic screening of the *WT1* sequence in 92 patients diagnosed with SFF (mostly NOA) revealed 1 missense substitution in exon 1 and no variation in the regions coding for the C-terminus of the protein. Additional 107 patients were tested for exons 1-6, which code for the N-terminal region of WT1 and overall two missense substitutions were identified in exons 1 -p.Pro130Leu- and 6 -p.Cys350Arg (Paper III Figure 3). The former likely hinders WT1 homodimerization through the self-association domain whereas p.Cys350Arg might interfere with the stability of the ZF domain given its close proximity. Interestingly, p.Cys350Arg was present in 3 patients with severe oligozoospermia (SOZ) and one with anorchia. This illustrates the spectrum of phenotypic expression of genetic defects at this locus and supports a model where a load of mildly deleterious variants represent risk factors for SFF, similarly to what was discussed for *DMRT1* variants, whereas highly damaging variants may result in more severe phenotypes. Genetic variation that does not change the sequence of the protein can be related to alterations in specific binding sites of regulatory proteins (Goymer 2007) and may affect specific pathways of germ cell maintenance. Also, 31 patients with unilateral or bilateral cryptorchidism were evaluated but no additional alterations were detected. Although the sample size needs to be extended, other trans-acting factors seem to play a role in gonadal development and function, portraying the multifactorial nature of male infertility phenotypes. These observations, together with the overall strikingly high conservation of the WT1 sequence in vertebrates (Paper III Figure 2), especially in the ZF domain, support the previously proposed model of haploinsufficiency for WT1 in both non-syndromic and more severe phenotypic manifestations depending on the genetic region compromised. A dominant-negative effect has also been proposed for WT1 mutations in Denys-Drash syndrome (Little and Wells 1997).

Collectively, the work reported in papers II and III suggests that: 1) genetic variants in *loci* involved in testicular development and spermatogenesis maintenance may have variable levels of penetrance that is manifested in terms of degrees of disease severity; 2) different types of genetic defects have variable phenotypic outcomes and need to be evaluated independently for each locus taking into account the predicted function of gene product in gonadal development and function; 3) SFF is a multifactorial disease associated with rare genetic defects and may result from a combinatory effect of mildly deleterious variation that primarily affects the regulation of gonadal function. Furthermore, the mode of action of different genes in the context of male infertility is likely related with the role they play in the pathways regulating gonadal development and spermatogenesis maintenance. Genetic defects in

transcription factors are expected to have a stronger phenotypic impact since it would influence many of their downstream target genes.

Determining causality of genetic variation in case-control studies of male infertility

Spermatogenesis is regulated by a delicate interplay of thousands of factors coded by many loci throughout the genome (Matzuk and Lamb 2008). It is therefore reasonable to expect that, at these loci, different types of genetic defects altering protein function and dose, disrupting protein-binding sites or interfering with mRNA splicing and translation can unbalance the regulatory network and compromise fertility in various ways. Typically, the search for disease causing genetic variation is mainly focused on missense substitutions given their potential to directly impact protein function. However, the work presented in this chapter emphasizes the potential role of both coding and non-coding regulatory variants in SFF, which can be more relevant than previously appreciated. The missense variants detected in the *WT1* coding sequence were localized in the self-association and transcriptional regulatory domains (Exon 1 and Exon 6, respectively; (Wang et al. 2013)). Instead of having a strong impact on protein function, since the main ZF domain remains intact, variation in these domains may interfere with *WT1* expression and/or specific pathways that are essential for the regulation of gonadal function but not development. Similarly, coding and non-coding variation with potential regulatory roles in *DMRT1* expression and splicing seems to be a risk factor for SFF.

The functional impact of regulatory variation is difficult to assess and mainly relies on *in silico* analyses, similar to those performed here (Papers II and III), and experimental validation of candidates. In fact, determining causality of genetic variants is a generalized issue in human genetics that led MacArthur and co-workers to propose guidelines for study design, analysis and report of disease-causing candidate variants (MacArthur *et al.* 2014). Large genome sequencing projects comprise a valuable repository of genetic diversity in healthy individuals of different genetic ancestries and can inform the evaluation of the deleterious effect of candidate variants. However, data from these projects revealed a strikingly high level of diversity within individual genomes where 2.3% of single nucleotide variants (SNVs) detected are predicted to affect protein function, ~95.7% of which are rare (Tennessen *et al.* 2012). Moreover, tolerance to variation varies amongst genes, with classes such as transcription factors showing an excess of rapidly evolving genes (Bustamante *et al.* 2005), and predicted

loss-of-function (LoF) variants are also detected in healthy individuals (MacArthur *et al.* 2012). Given the complex multifactorial genetic architecture underlying phenotypes of SFF, establishing the causality of rare genetic variants at specific loci is particularly challenging in cohorts of infertile men, more so when considering the potential role of mildly deleterious regulatory variants. Family-based studies would allow to assess co-segregation of candidate variants with disease status (Samocha *et al.* 2014), but these are seldom available in cases of human male infertility (See paper I). Other specific issues involve access to testicular tissue, lack of a reliable *in vitro* model for spermatogenesis and phenotype classification. In male infertility cases, while some can exhibit multiple abnormal sperm parameters, such as defects in sperm count and sperm morphology, others may represent intermediate phenotypes of disease progression (See Paper I). Mixed cohorts comprising patients with different disease etiologies would further compromise the already challenging detection of disease-causing variation. For instance, the primary assessment of one patient carrying *WT1* variant in exon 6 revealed a NOA phenotype which was later redefined as SOZ upon a second attempt of retrieving sperm from the ejaculate. This demonstrates the difficulty of establishing cohorts with well-defined phenotypes and highlights the importance of clinician-researcher communication and patient follow-up.

In this chapter, to evaluate potential deleterious effects of variants detected in patient cohorts, several premises were followed, in accordance to (MacArthur *et al.* 2014): 1) quantitative estimation of variant frequency and assessment of statistical significance in patients VS control populations; 2) evaluation of sequence conservation and 3) predicted effect on function. The latter was assessed *in silico* by different methods according to the type of variation. Effect of missense substitutions in *WT1* were evaluated considering their sequence context (protein domain or binding-sites) and deleteriousness was assessed using two algorithms (Polyphen-2 and SIFT). Synonymous substitutions and non-coding variants in *DMRT1* overrepresented in patients were analyzed for putative regulatory effects on gene expression and mRNA splicing using two different available bioinformatics tools for each analysis (mRNA splicing: Human Splicing Finder and BDGP: Splice Site Prediction by Neural Network; TFBSs in the promoter: MatInspector and TFSEARCH). As comparative sequence analysis is a powerful source of information regarding deleteriousness (Cooper and Shendure 2011), evolutionary sequence conservation was evaluated for both protein-coding and non-coding variation. These approaches allowed the prioritization of candidate variants in both *WT1* and *DMRT1* sequences for further experimental validation, which is ultimately required to establish disease causality (MacArthur *et al.* 2014).

Determination of the functional impact of new genetic variants in male infertility can be facilitated by integration of more detailed clinical phenotypes and genetic data from a larger number of patients. Indeed, sample size is a common limitation in case-control studies but essential to associate rare variants with complex traits (Tennessen *et al.* 2012). Applying population genetics principles and sound statistical approaches controlling for population structure and patient's ancestry should also be considered. In this regard, statistical models to estimate mutation rates at specific loci with single nucleotide resolution (Schaibley *et al.* 2013) and to predict the cumulative effects of patient-specific variants (Lopes *et al.* 2013) should prove extremely useful. Furthermore, candidate-gene studies in male infertility are limited by the fact that they fail to capture variation in other potentially relevant trans-acting factors. Nonetheless, they are useful hypothesis-driven follow-up approaches that allow the identification of novel variation further implicating those loci in disease, as exemplified by the works in papers II and III. *DMRT1* and *WT1* screens provided a global picture of the genetic diversity at these loci as well as a population-specific appraisal of potentially deleterious rare variants in men with SFF as an isolated phenotype or in the context of other gonadal abnormalities. The results highlighted the relevance of different types of variation in phenotype expression. A comprehensive functional annotation of the genome, in view with the guidelines discussed above, should contribute to strengthen predictions of pathogenicity for different variants and to a better understanding of their relevance in male infertility.

Literature Cited

Anand, A., Patel, M., Lalremruata, A., Singh, A.P., Agrawal, R., Singh, L. and Aggarwal, R.K. (2008). Multiple alternative splicing of Dmrt1 during gonadogenesis in Indian mugger, a species exhibiting temperature-dependent sex determination. *Gene,* **425**: 1-2, 56-63.

Barbaro, M., Balsamo, A., Anderlid, B.M., Myhre, A.G., Gennari, M., Nicoletti, A., Pittalis, M.C., Oscarson, M. and Wedell, A. (2009). Characterization of deletions at 9p affecting the candidate regions for sex reversal and deletion 9p syndrome by MLPA. *Eur J Hum Genet,* **17**: 11, 1439-1447.

Barrionuevo, F.J., Burgos, M., Scherer, G. and Jimenez, R. (2012). Genes promoting and disturbing testis development. *Histol Histopathol,* **27**: 11, 1361-1383.

Bratuś, A. (2012) DM Domain Genes : Sexual and Somatic Development During Vertebrate Embryogenesis. sas978-953-51-0466-7.

Bustamante, C.D., Fledel-Alon, A., Williamson, S., Nielsen, R., Hubisz, M.T., Glanowski, S., Tanenbaum, D.M., White, T.J., Sninsky, J.J., Hernandez, R.D., Civello, D., Adams, M.D., Cargill, M. and Clark, A.G. (2005). Natural selection on protein-coding genes in the human genome. *Nature,* **437**: 7062, 1153-1157.

Calvel, P., Rolland, A.D., Jegou, B. and Pineau, C. (2010). Testicular postgenomics: targeting the regulation of spermatogenesis. *Philos Trans R Soc Lond B Biol Sci,* **365**: 1546, 1481-1500.

Chau, Y.Y., Brownstein, D., Mjoseng, H., Lee, W.C., Buza-Vidas, N., Nerlov, C., Jacobsen, S.E., Perry, P., Berry, R., Thornburn, A., Sexton, D., Morton, N., Hohenstein, P., Freyer, E., Samuel, K., van't Hof, R. and Hastie, N. (2011). Acute multiple organ failure in adult mice deleted for the developmental regulator Wt1. *PLoS Genet,* **7**: 12, e1002404.

Cheng, H.H., Ying, M., Tian, Y.H., Guo, Y., McElreavey, K. and Zhou, R.J. (2006). Transcriptional diversity of DMRT1 (dsx- and mab3-related transcription factor 1) in human testis. *Cell Res,* **16**: 4, 389-393.

Cooper, G.M. and Shendure, J. (2011). Needles in stacks of needles: finding disease-causal variants in a wealth of genomic data. *Nat Rev Genet,* **12**: 9, 628-640.

Eggers, S. and Sinclair, A. (2012). Mammalian sex determination-insights from humans and mice. *Chromosome Res,* **20**: 1, 215-238.

Erdman, S.E. and Burtis, K.C. (1993). The Drosophila doublesex proteins share a novel zinc finger related DNA binding domain. *EMBO J,* **12**: 2, 527-535.

Gessler, M., Konig, A. and Bruns, G.A. (1992). The genomic organization and expression of the WT1 gene. *Genomics,* **12**: 4, 807-813.

Goymer, P. (2007). Synonymous mutations break their silence. *Nature Reviews Genetics,* **8**: 2, 92-92.

Haber, D.A., Sohn, R.L., Buckler, A.J., Pelletier, J., Call, K.M. and Housman, D.E. (1991). Alternative splicing and genomic structure of the Wilms tumor gene WT1. *Proc Natl Acad Sci U S A,* **88**: 21, 9618-9622.

Hammes, A., Guo, J.K., Lutsch, G., Leheste, J.R., Landrock, D., Ziegler, U., Gubler, M.C. and Schedl, A. (2001). Two splice variants of the Wilms' tumor 1 gene have distinct functions during sex determination and nephron formation. *Cell,* **106**: 3, 319-329.

Herpin, A., Braasch, I., Kraeussling, M., Schmidt, C., Thoma, E.C., Nakamura, S., Tanaka, M. and Schartl, M. (2010). Transcriptional rewiring of the sex determining dmrt1 gene duplicate by transposable elements. *PLoS Genet,* **6**: 2, e1000844.

Holmes, G., Boterashvili, S., English, M., Wainwright, B., Licht, J. and Little, M. (1997). Two N-terminal self-association domains are required for the dominant negative transcriptional activity of WT1 Denys-Drash mutant proteins. *Biochem Biophys Res Commun,* **233**: 3, 723-728.

Hossain, A. and Saunders, G.F. (2001). The human sex-determining gene SRY is a direct target of WT1. *J Biol Chem,* **276**: 20, 16817-16823.

Huff, V. (2011). Wilms' tumours: about tumour suppressor genes, an oncogene and a chameleon gene. *Nat Rev Cancer,* **11**: 2, 111-121.

Jorgensen, A., Nielsen, J.E., Blomberg Jensen, M., Graem, N. and Rajpert-De Meyts, E. (2012). Analysis of meiosis regulators in human gonads: a sexually dimorphic spatio-temporal expression pattern suggests involvement of DMRT1 in meiotic entry. *Mol Hum Reprod,* **18**: 11, 523-534.

Kim, S., Bardwell, V.J. and Zarkower, D. (2007). Cell type-autonomous and non-autonomous requirements for Dmrt1 in postnatal testis differentiation. *Dev Biol,* **307**: 2, 314-327.

Kohler, B., Biebermann, H., Friedsam, V., Gellermann, J., Maier, R.F., Pohl, M., Wieacker, P., Hiort, O., Gruters, A. and Krude, H. (2011). Analysis of the Wilms' tumor suppressor gene (WT1) in patients 46,XY disorders of sex development. *J Clin Endocrinol Metab,* **96**: 7, E1131-1136.

Little, M. and Wells, C. (1997). A clinical overview of WT1 gene mutations. *Hum Mutat,* **9**: 3, 209-225.

Lopes, A.M., Aston, K.I., Thompson, E., Carvalho, F., Goncalves, J.*, et al.* (2013). Human Spermatogenic Failure Purges Deleterious Mutation Load from the Autosomes and Both Sex Chromosomes, including the Gene DMRT1. *Plos Genetics,* **9**: 3,

Lu, H., Huang, X., Zhang, L., Guo, Y., Cheng, H. and Zhou, R. (2007). Multiple alternative splicing of mouse Dmrt1 during gonadal differentiation. *Biochem Biophys Res Commun,* **352**: 3, 630-634.

MacArthur, D.G., Balasubramanian, S., Frankish, A., Huang, N., Morris, J.*, et al.* (2012). A systematic survey of loss-of-function variants in human protein-coding genes. *Science,* **335**: 6070, 823-828.

MacArthur, D.G., Manolio, T.A., Dimmock, D.P., Rehm, H.L., Shendure, J.*, et al.* (2014). Guidelines for investigating causality of sequence variants in human disease. *Nature,* **508**: 7497, 469-476.

Machado, A.Z., da Silva, T.E., Costa, E.M.F., dos Santos, M.G., Nishi, M.Y., Brito, V.N., Mendonca, B.B. and Domenice, S. (2012). Absence of inactivating mutations and deletions in the DMRT1 and FGF9 genes in a large cohort of 46,XY patients with gonadal dysgenesis. *European Journal of Medical Genetics,* **55**: 12, 690-694.

Matson, C.K., Murphy, M.W., Sarver, A.L., Griswold, M.D., Bardwell, V.J. and Zarkower, D. (2011). DMRT1 prevents female reprogramming in the postnatal mammalian testis. *Nature,* **476**: 7358, 101-+.

Matzuk, M.M. and Lamb, D.J. (2008). The biology of infertility: research advances and clinical challenges. *Nature Medicine,* **14**: 11, 1197-1213.

Moniot, B., Berta, P., Scherer, G., Sudbeck, P. and Poulat, F. (2000). Male specific expression suggests role of DMRT1 in human sex determination. *Mechanisms of Development,* **91**: 1-2, 323-325.

Murphy, M.W., Sarver, A.L., Rice, D., Hatzi, K., Ye, K., Melnick, A., Heckert, L.L., Zarkower, D. and Bardwell, V.J. (2010). Genome-wide analysis of DNA binding and transcriptional regulation by the mammalian Doublesex homolog DMRT1 in the juvenile testis. *Proceedings of the National Academy of Sciences of the United States of America,* **107**: 30, 13360-13365.

Parenti, R., Salvatorelli, L., Musumeci, G., Parenti, C., Giorlandino, A., Motta, F. and Magro, G. (2015). Wilms' tumor 1 (WT1) protein expression in human developing tissues. *Acta Histochem,* **117**: 4-5, 386-396.

Practice Committee of American Society for Reproductive Medicine in collaboration with Society for Male Reproduction and Urology (2008). Evaluation of the azoospermic male. *Fertil Steril,* **90**: 5 Suppl, S74-77.

Quinonez, S.C., Park, J.M., Rabah, R., Owens, K.M., Yashar, B.M., Glover, T.W. and Keegan, C.E. (2013). 9p partial monosomy and disorders of sex development: review and postulation of a pathogenetic mechanism. *Am J Med Genet A,* **161A**: 8, 1882-1896.

Raymond, C.S., Murphy, M.W., O'Sullivan, M.G., Bardwell, V.J. and Zarkower, D. (2000). Dmrt1, a gene related to worm and fly sexual regulators, is required for mammalian testis differentiation. *Genes & Development,* **14**: 20, 2587-2595.

Raymond, C.S., Shamu, C.E., Shen, M.M., Seifert, K.J., Hirsch, B., Hodgkin, J. and Zarkower, D. (1998). Evidence for evolutionary conservation of sex-determining genes. *Nature,* **391**: 6668, 691-695.

Rey, R.A. and Grinspon, R.P. (2011). Normal male sexual differentiation and aetiology of disorders of sex development. *Best Pract Res Clin Endocrinol Metab,* **25**: 2, 221-238.

Royer-Pokora, B., Beier, M., Henzler, M., Alam, R., Schumacher, V., Weirich, A. and Huff, V. (2004). Twenty-four new cases of WT1 germline mutations and review of the literature: genotype/phenotype correlations for Wilms tumor development. *Am J Med Genet A,* **127A**: 3, 249-257.

Samocha, K.E., Robinson, E.B., Sanders, S.J., Stevens, C., Sabo, A.*, et al.* (2014). A framework for the interpretation of de novo mutation in human disease. *Nat Genet,* **46**: 9, 944-950.

Schaibley, V.M., Zawistowski, M., Wegmann, D., Ehm, M.G., Nelson, M.R., St Jean, P.L., Abecasis, G.R., Novembre, J., Zollner, S. and Li, J.Z. (2013). The influence of genomic context on mutation patterns in the human genome inferred from rare variants. *Genome Res,* **23**: 12, 1974-1984.

Schieck, M., Schouten, J.P., Michel, S., Suttner, K., Toncheva, A.A.*, et al.* (2016). Doublesex and mab-3 related transcription factor 1 (DMRT1) is a sex-specific genetic determinant of childhood-onset asthma and is expressed in testis and macrophages. *J Allergy Clin Immunol,*

Seabra, C.M. (2012) Rare structural variants in severe spermatogenic impairment. Universidade de Aveiro, Aveiro, Portugal,

Seabra, C.M., Quental, S., Neto, A.P., Carvalho, F., Goncalves, J., Oliveira, J.P., Fernandes, S., Sousa, M., Barros, A., Amorim, A. and Lopes, A.M. (2014). A novel Alu-mediated microdeletion at 11p13 removes WT1 in a patient with cryptorchidism and azoospermia. *Reprod Biomed Online,* **29**: 3, 388-391.

Strachan, T. and Read, A. (2011a). Genetic Mapping of Mendelian Characters *in*: Human Molecular Genetics, Garland Science, 14, 441-466; 978-0-815-34149-9.

Strachan, T. and Read, A. (2011b). Mapping Genes Confering Susceptibility to Complex Diseases *in*: Human Molecular Genetics, Garland Science, 15, 467-496; 978-0-815-34149-9.

Tannour-Louet, M., Han, S., Corbett, S.T., Louet, J.F., Yatsenko, S., Meyers, L., Shaw, C.A., Kang, S.H.L., Cheung, S.W. and Lamb, D.J. (2010). Identification of De Novo Copy Number Variants Associated with Human Disorders of Sexual Development. *Plos One,* **5**: 10, e15392.

Tennessen, J.A., Bigham, A.W., O'Connor, T.D., Fu, W., Kenny, E.E.*, et al.* (2012). Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science,* **337**: 6090, 64-69.

Tewes, A.C., Ledig, S., Tuttelmann, F., Kliesch, S. and Wieacker, P. (2014). DMRT1 mutations are rarely associated with male infertility. *Fertil Steril,* **102**: 3, 816-820 e813.

Toska, E. and Roberts, S.G. (2014). Mechanisms of transcriptional regulation by WT1 (Wilms' tumour 1). *Biochem J,* **461**: 1, 15-32.

Wang, X.N., Li, Z.S., Ren, Y., Jiang, T., Wang, Y.Q.*, et al.* (2013). The Wilms tumor gene, Wt1, is critical for mouse spermatogenesis via regulation of sertoli cell polarity and is associated with non-obstructive azoospermia in humans. *PLoS Genet,* **9**: 8, e1003645.

Wilhelm, D. and Englert, C. (2002). The Wilms tumor suppressor WT1 regulates early gonad development by activation of Sf1. *Genes Dev,* **16**: 14, 1839-1851.

Xu, S., Han, J.C., Morales, A., Menzie, C.M., Williams, K. and Fan, Y.S. (2008). Characterization of 11p14-p12 deletion in WAGR syndrome by array CGH for identifying genes contributing to mental retardation and autism. *Cytogenet Genome Res,* **122**: 2, 181-187.

Zarkower, D. (2013). DMRT Genes in Vertebrate Gametogenesis. *Gametogenesis,* **102**: 327-356.

Zhao, L., Svingen, T., Ng, E.T. and Koopman, P. (2015). Female-to-male sex reversal in mice caused by transgenic overexpression of Dmrt1. *Development,* **142**: 6, 1083-1088.

CHAPTER 2

Integrative functional genomics in the context of male infertility: developing a new approach

Male infertility is a complex disorder comprising a spectrum of different phenotypes arising from alterations in any of the panoply of genes and regulatory networks controlling normal progression of spermatogenesis. In this chapter, I aimed to better understand the underlying biology of spermatogenesis, focusing on the processes of gene expression and protein translation. For that I explored and developed methods that, while addressing the cellular heterogeneity of the testis, would allow to gather high-throughput data for functional genomic studies with an evolutionary perspective on spermatogenesis. We expect that these studies will ultimately allow to gain insights into the functional relevance of regulatory and coding genetic variants found in the context of male infertility.

Introduction

The evolution of multicellularity has brought functional specialization of tissues and allowed for the development of unique traits and systems as an organism's response to external factors. There is an intrinsic cellular heterogeneity within a tissue's architecture, comprising layers of intricate inter- and intracellular regulatory mechanisms. In adult organs, tissue complexity is related to highly specialized functions and an array of greatly diverse and differentiated cells. For instance, testicular germ cells display stage-specific characteristics associated with unique molecular events that are crucial for spermatogenesis (meiosis, chromatin remodelling, repackaging and transcriptional reprogramming). Control of these processes involves a tight coordination of juxtacrine, paracrine and endocrine factors and highly dynamic gene expression and protein translation regulation, including the generation of an unusual amount of tissue-specific isoforms (Calvel *et al.* 2010). These features are maintained by over 30 distinct cell types/states (Rodriguez-Casuriaga *et al.* 2013) and together make the testis, alongside with brain, one of the most complex tissues in the body. Thus, understanding what renders the uniqueness of specific germ cell types provides invaluable clues about their role and importance for tissue homeostasis. Despite recent improvements (Q. Zhou *et al.* 2016), the lack of a solid standard system for *in vitro* culture of male germ cells has been placing greater emphasis on the quality of *ex vivo* studies (Wistuba *et al.* 2007; Chocu *et al.* 2012) and further complicates studies addressing specific cell types. This raises two relevant issues: 1- researchers need robust and efficient methods to discriminate and isolate different cell-types; and 2- the amount of working material is drastically and proportionally reduced in function of the frequency of those specific cell-types within the tissue.

Isolation of male germ cells

Much like puzzle pieces, the molecular signatures of distinct germ cell types portrait the cellular programing of specific developmental stages. Individual profiles of such populations are, therefore, powerful sources to tackle the strict molecular program driving progression of spermatogenesis. This has been motivating researchers to develop and optimize techniques for the isolation of different spermatogenic cells. In general, the most commonly used techniques rely on different cellular attributes such as size, density, shape and/or chromatin variations to discriminate cells in different developmental stages. StaPut gravity sedimentation (Lam *et al.* 1970; Romrell *et al.* 1976; Bryant *et al.* 2013) and elutriation (Meistrich 1977; Chang *et al.* 2011) make use of bovine serum albumin (BSA) or Percoll gradients, respectively, to enrich for fractions based on variations in cell size and density during spermatogenesis. Higher

purities can be achieved by cell sorting, using either makers for specific germ cell-types (Magnetic-Activated Cell Sorting; MACS;Gassei *et al.* (2010); He *et al.* (2012); Grunewald and Paasch (2013)) or vital fluorochromes that discriminate developmental stages based on chromatin structure and amount (Fluorescence-Activated Cell Sorting; FACS; Bastos *et al.* (2005); Geisinger and Rodriguez-Casuriaga (2010); Getun *et al.* (2011); Gaysinskaya *et al.* (2014); Gaysinskaya and Bortvin (2015)). The most widely applied method for mouse, perhaps given its simplicity, has been the collection of testis samples at specific days post-partum (dpp), timed for the first appearance of the distinct germ cell-types during the first wave of spermatogenesis (Schultz *et al.* 2003; Shima *et al.* 2004; Margolin *et al.* 2014). However, pre-pubertal testis have been shown to be molecularly distinct from adult testis (Yoshida *et al.* 2006) and therefore provide insights into the coordinated development of testis architecture rather than the maintenance of testicular biology and formation of spermatozoa.

When evaluating/choosing a method for enrichment of specific cell types, be it from testis or any other tissue, certain characteristics must be taken into account: i) model organism in which it has been developed/applied to; ii) yield of cells/population/biological sample; iii) resolution of the method, i.e., how many different cell types it discriminates; iv) impact of the procedure on cellular biology (toxicity of reagents used, stress conditions, length, etc.); v) labor intensiveness and vi) overall cost. Table 1 summarizes the main pros and cons of each method, providing a generalized idea of their strengths and limitations in the isolation of male germ cells. Obviously, compromises will most likely have to be made and the choice of method will ultimately rely on the goal of the study, researcher's skills and access to the required materials. Given that our aim was to investigate the molecular complexity of spermatogenesis with a stage-wise functional genomics approach, FACS seemed the most suitable technique, as it provides the highest resolution of germ cell types and good yield and purity.

Basic principles of Fluorescence-Activated Cell Sorting (FACS)

The creation and commercialization of the first fluorescence-activated cell sorters dates back to almost 50 years ago (Bonner *et al.* 1972; Herzenberg *et al.* 1976). FACS has since benefitted from technological advances in the fields of optics, electronics and data analysis allied to the development of monoclonal antibodies and new fluorochromes. Specifically, the technical milestone of polychromatic flow cytometry (simultaneous detection of five or more markers) led to the generation of multi-parameter flow cytometers, currently allowing the

detection of 18 colors at a rate of 1000 cells/s (Reviewed in Bendall *et al.* (2012); Galbraith (2012)). This makes flow cytometry unique in its ability to identify and characterize populations of cells from heterogeneous mixtures by single-cell 'deep profiling' of multiple biologic signatures. The applications are immense, ranging from DNA analysis (cell cycle, ploidy, cell proliferation, etc.), to cell counting, GFP expression analysis, intracellular β-galactosidase reporter gene assays (FACS Gal), individual clone sorting during hybridoma development and immunophenotyping (Reviewed in Herzenberg *et al.* (2002); Ormerod (2008)). The latter is

Table 1. Different methods for germ cell enrichment.

Method	Starting material	Yield (cells/pop)	Purity	Resolution	Time	Cost	Main advantage	Main disadvantage	Reference
STA-PUT	22 testes	10^8	90%	Meiotic+Somatic; rSpd; eSpd	3-4h	$$	Large yields	Low resolution - Mixed Spg, Spc and somatic cells	Bryant *et al.* (2013)
Elutriation	2 testes	10^7	80-95%	Spc I; rSpd;eSpd; Leydig; Sertoli	3-4h	$$	Collection of somatic and germ cells simultaneously	Low resolution	Chang *et al.* (2011)
MACS	1 testis	Dependent on marker	>95%	Spg; Spz	5-10h	$$$	High purity	Specific markers	He *et al. (* 2012); Grunewald *et al.* (2013)
FACS	1 testis	0.5- 13 x 10^6	75-95%	Spg; PreL; Spc I (4 subtypes); Spc II; Spd (4 subtypes)	2-4h	$$$	Collects 4 populations at a time Highest resolution of Spc and Spd	High-tech equipment and skilled technician	Lima *et al.* (paper IV); Simard (2016); Gaysinskaya (2014)
First wave	n= # stages under study	bulk tissue	n.a.	5 time points (Spg; Spc; rSpd;eSpd; Spz)	n.a.	$	No cell separation required	Testis ontogenesis different from adult spermatogenesis	Margolin *et al. (2014)* ; Yoshida *et al.* (2006)

responsible for stablishing FACS as the 'golden standard' technique to address the complexity of immune cells (Reviewed in Chattopadhyay and Roederer (2012)). But how are cells distinguished and then separated based on these different parameters?

Flow cytometry analyses are based on the detection of differential light patterns emerging from laser beam irradiation of individual cells. Figure 4 illustrates an overview of a cell sorter. Essentially, as a cell passes through the laser it will reflect/scatter light at all angles, which will be quantified by a detector that converts light intensity into voltage pulses. The amount of light that scatters in the forward direction (forward scatter; FS) is proportional to the size of the cell, whereas granularity and structural complexity inside the cell influence light scattering to the sides (side scatter; SS). Signals from the detectors are converted into digital data that can be analyzed statistically by flow cytometry software. 2D scatter plots obtained as a function of FS and SS allow to distinguish populations of cells based on these cellular features. Similarly, when cells are stained with fluorophores, the fluorescence signal is directed through a series of filters and mirrors to the appropriate detectors, where it is translated into a voltage pulse proportional to the amount of fluorescence emitted. Hence, users can gate and

sort populations based on the analysis of FS, SS, presence/absence of specific markers or as a function of fluorescence signals of the fluorophores. Mechanically, this is achieved by a coordinated action involving a fluidics system, lasers, optics, detectors, electronics and a peripheral computer system. The most common method of sorting cells is by electrostatic deflection of charged droplets. In the flow chamber, as the sample is injected in the sheath fluid (water or saline buffer), its own stream is compressed and the diameter narrowed to deliver single cells to the interrogation point, where the laser and the sample intersect. Vertical vibration of the flow chamber by a piezoelectric transducer causes the fluid to break up into droplets. When a cell of interest is detected at the interrogation point and is inside the droplet currently being formed, the flow stream receives a positive or negative charge (50-150V). The charged droplets are then deflected as they pass through a pair of charged plates (± 5000 V), allowing the collection of identically charged droplets. Detailed information about the technique, recent developments and general applications can be found elsewhere (Herzenberg et al. 2002; Ormerod 2008; Bendall et al. 2012; Galbraith 2012).

Flow cytometry of spermatogenic cells

Over the past 40 years, flow cytometry has been extensively applied in the field of male reproductive biology (Reviewed in Geisinger and Rodriguez-Casuriaga (2010)). Direct clinical applications include: i) gamete quality assessment by means of sperm viability, acrosomal integrity and sperm motility (De Iuliis *et al.* 2009; Kordan *et al.* 2013; Robles and Martinez-Pastor 2013; Jenkins *et al.* 2015); ii) diagnosis of male infertility (Omran *et al.* 2013; Belloc *et al.* 2014); iii) evaluation of animal fertilizing ability (Druart *et al.* 2009); and iv) sex-sorting of X/Y chromosome-bearing sperm for offspring gender control in veterinary science (Reviewed in Garner (2006)). As a powerful tool in basic biology of spermatogenesis, flow cytometry has been useful to identify spermatogonial stem cell (SSC) markers (Kim *et al.* 2013; Rafeeqi and Kaul 2013), to characterize and isolate SSC populations (Shinohara *et al.* 2000; Hermann *et al.* 2009) and/or embryonic stem cell (ESC)-derived primordial germ cells (PGCLCs) for in vitro culture (Ryu *et al.* 2005; Q. Zhou *et al.* 2016) and transplantation purposes (Falciatori *et al.* 2004; Lassalle *et al.* 2004), to quantify apoptotic testicular germ cells (Krishnamurthy *et al.* 1998; K. W. Zhou *et al.* 2009) and to assess testicular heterogeneity (Mays-Hoopes *et al.* 1995;

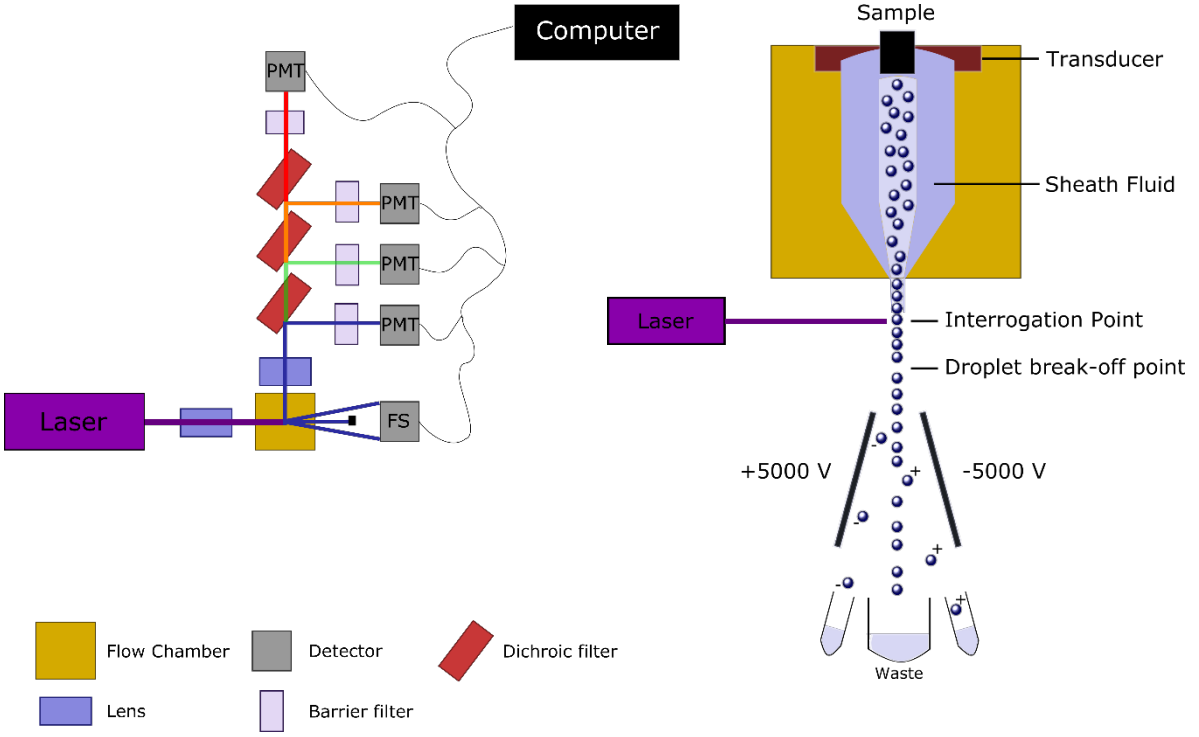


Figure 4. Overview of a cell sorter: components and sorting mechanism.

A typical cell sorter is equipped with a fluidics system and a series of lasers, optics, detectors, and electronics arranged so that variations in light signals can be detected, measured and delivered to a peripheral computer system (left). The most common method for cell sorting is by electrostatic deflection of charged droplets, depicted in the right image. More detailed information can be found in the text. Adapted from Ormerod (2008).

Rodriguez-Casuriaga *et al.* 2011) and ontogenesis (Rotgers *et al.* 2015) in different species, among others (Reviewed in Geisinger and Rodriguez-Casuriaga (2010)).

The nature of spermatogenesis progression, with cells in the different stages varying widely in DNA content, chromatin structure, size and shape of their nuclei, allows for a straightforward identification of different cell populations by combining light scattering and DNA staining with vital dyes (Grogan *et al.* 1981; Mays-Hoopes *et al.* 1995). Other differentiation-associated parameters detected by flow cytometry, such as RNA content (D. P. Evenson and Melamed 1983) and changes in mitochondria amount (Petit *et al.* 1995) and activity (Cordelli *et al.* 2005; De Iuliis *et al.* 2009), have also been considered. For these purposes, several fluorescent dyes with different excitation and emission spectra, permeability and binding

properties have been used and evaluated in respect of stainability, toxicity and applicability (D. Evenson *et al.* 1986; Geisinger and Rodriguez-Casuriaga 2010). As cellular integrity needs to be preserved, studies using sorted testicular populations as individual biological samples for downstream procedures require dyes capable of crossing intact cell membranes.

Hoechst-33342 (Ho) is a cell-permeant vital dye that has been used for decades in flow cytometry analysis of testicular cells, alone or in combination with the non-permeant Propidium Iodine (PI) to define living versus dead cells (Grogan *et al.* 1981; Bastos *et al.* 2005; Gaysinskaya *et al.* 2014; Gaysinskaya and Bortvin 2015). It binds preferentially to poly(d[AT]) sequences in the minor groove of DNA and emits blue fluorescence proportional to DNA content when excited with UV light ((Bastos *et al.* 2005); Fig. 5). At higher ratios, secondary binding takes place reflecting conformational changes in DNA (Watson *et al.* 1985) that emit far red fluorescence (Goodell *et al.* 1996). Exceptionally, as a result of BCRP1-dependent dye efflux, which is switched off after the spermatogonial stages, Ho stained spermatogonial stem cells show a unique fluorescence pattern and represent a side population (Bastos *et al.* 2005). Therefore, measuring Ho intensity as a function of blue and red fluorescence is representative of 3 cellular properties: ploidy, chromatin structure/accessibility, and dye efflux caused by ABC transporter activity (Goodell *et al.* 1996; Falciatori *et al.* 2004; Lassalle *et al.* 2004; Bastos *et al.* 2005; Gaysinskaya *et al.* 2014). Importantly, flow cytometry with Ho staining (Ho-FACS) has been shown to allow the isolation of 9 different germ cell types: spermatogonia, primary spermatocytes in different meiotic prophase stages (preleptotene, leptotene, zygotene, pachytene and diplotene), secondary spermatocytes and post-meiotic round and elongating spermatids (Lassalle *et al.* 2004; Bastos *et al.* 2005; Shimizu *et al.* 2006; Getun *et al.* 2011; Gaysinskaya *et al.* 2014; Gaysinskaya and Bortvin 2015).

Remarkably, flow cytometry analysis of different mammalian species, such as mouse, hamster, pig, cat and several primates including men, has revealed similar profiles in terms of DNA ploidy/stainability (Reviewed in Geisinger and Rodriguez-Casuriaga (2010)). Despite the enormous contribution to the optimization of flow cytometry analysis of testicular heterogeneity in many individual species, these studies have used different dyes, staining protocols and flow cytometry parameters for analysis. Importantly, the high resolution power of Ho-FACS providing the isolation of 9 germ-cell types has only been optimized in mouse. Therefore, comparative studies in male reproductive biology are hindered by the lack of a standardized transversal flow cytometry procedure for isolation of specific male germ cells from different species.

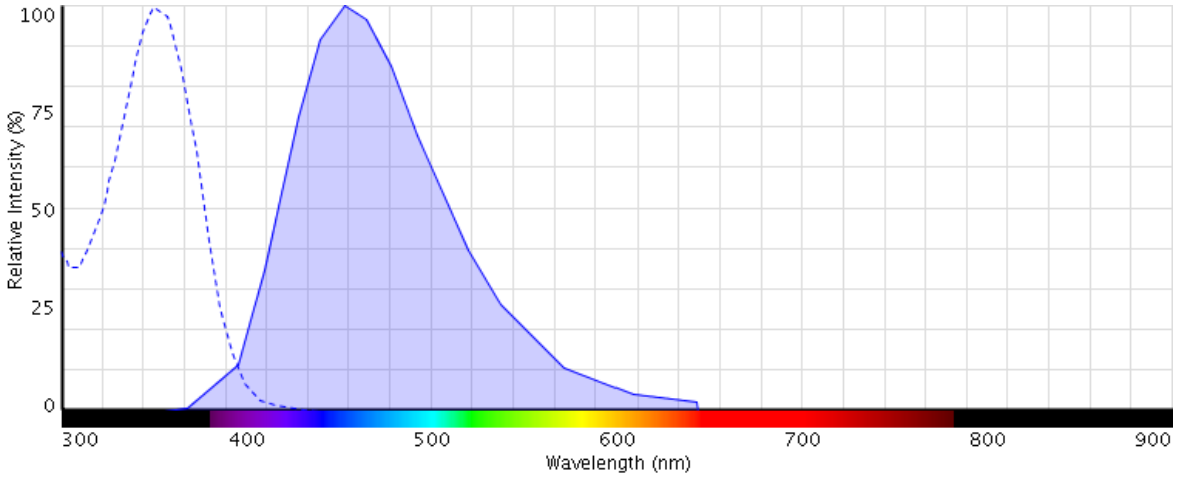


Figure 5. Hoechst-33342 excitation and emission spectra.

Hoechst-33342 is excited (350 nm) by a UV laser and emits a peak of blue fluorescence (461 nm), proportional to the amount of DNA present in the cell. However, it also emits the far red region, which allows for the detection of different states of chromatin conformation. Image retrieved on 6-3-2016 from https://www.thermofisher.com/us/en/home/life-science/cell-analysis/labeling-chemistry/fluorescence-spectraviewer.html?ICID=svtool&UID=1398dna_3.

The complexity of spermatogenesis and consequences for the omics approaches

As genomes are essentially invariant within an organism, the cellular plasticity required for responses to both internal and external stimuli relies on this dynamic multilayered gene and protein expression system. Consequently, the biological relevance of genetic diversity catalogued in high-throughput whole-genome sequencing data is often challenging to assess and requires contextual information provided by functional annotation of genomes. Indeed, as stated above, this has been one of the major issues in genome-wide association studies (GWAS) of male infertility (Reviewed in Carrell *et al.* (2016)). This awareness and the technological advances allowing the application of genome technologies to capture gene and protein expression on a global scale drove the emergence of the recent field of functional genomics. Essentially, this thriving branch of genomics addresses the intricate levels regulating cell behavior by cataloguing various components of the regulome, transcriptome and proteome in an attempt to help decipher causal genotype - phenotype relationships.

When looking at germ cells in different developmental stages, this regulatory circuitry reaches a whole new level of complexity. Spermatogenic cells in different stages cannot be seen as individual entities since they represent intermediate steps of a progressive developmental process. This concept is clearly exemplified by the high proportion of mRNAs that are transcribed at one stage, stored in RNA granules, and translated only at a later point (Iguchi *et al.* 2006; Gan *et al.* 2013). Understanding the molecular mechanisms underlying spermatogenesis progression is therefore extremely challenging and will require the integrative analysis of different regulatory layers.

Postgenomic studies of spermatogenesis

Considering the cellular heterogeneity of the testis and the intricate coordinated molecular events controlling male gamete development, the collection of functional data and build-up of molecular regulatory networks seems to be the ideal framework towards a better understanding of this complex biological system. In this regard, several insightful high throughput studies have already begun to unravel the testicular transcriptome and proteome profiles of different species (Reviewed in Calvel *et al.* (2010); Com *et al.* (2014)). The strategies adopted by these studies led to either systematic characterization or evaluation of differential expression of gene and proteins of the testicular tissue in different species. For instance, systematic profiling of the ontogenic/whole testis provided extensive catalogues of the testicular

transcript and protein repertoires that allowed the identification of testis-specific isoforms and genes required for normal testis development and/or spermatogenesis progression. However, this approach remains highly descriptive and is limited by the fact that it lacks resolution about the spatial and temporal regulation during spermatogenesis. To surpass this issue, some studies made use of the variety of available methods for isolation of male germ cells (summarized above) to generate gene and protein expression profiles of specific germ cell types. All these studies generated extensive and comprehensive catalogues of testicular functional data, publicly available in online repositories for transcriptome (See Mooney and McWeeney (2014)) and Proteome (See T. Chen *et al.* (2015); Perez-Riverol *et al.* (2015)) data. Examples include the GermOnline (http://www.germonline.org/ ; Lardenois *et al.* (2010)), a database specific for high-throughput expression data relevant for germline development and the Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/), or the ProteomeXchange (http://www.proteomexchange.org/) and Proteomics Identifications databases (PRIDE; http://www.ebi.ac.uk/pride) for proteome datasets. Data mining of such large high-throughput datasets led to the development of many creative strategies and computational methods to provide meaningful functional annotations and enhance our overall understanding of the cellular events involved. One interesting approach employed was the cross-tissue and/or cross-species comparison of transcriptome data to identify evolutionarily-conserved and testis-specific genes. Based on the similarity of expression profiles, Chalmel and co-workers (Chalmel *et al.* 2007b) identified 888 orthologous genes as constituents of the core expression program conserved between human and rodents. Also, differential expression measured in 17 somatic control tissues and the testis, and/or germ cells *versus* testicular somatic cells, revealed a myriad of potential testis-specific transcripts that were mainly implicated in germ cell development given the reduced levels found in spermatogonia and Sertoli cells. Similarly, in a study by Soumillon and co-workers (Soumillon *et al.* 2013), a comparison with 5 somatic tissues across 5 different species representative of primates, rodents and birds, revealed an intricate testicular transcriptome with complex alternative splicing patterns, a pronounced diversity for noncoding transcripts, and the largest numbers of transcribed intergenic elements that mainly derived from meiotic spermatocytes and postmeiotic round spermatids. However, and not undervaluing their significant contributions, these studies fail to grasp the connections between co-expressed genes or proteins that ultimately explain their roles in spermatogenesis progression. One common way to try to unveil the network of molecular interactions is to evaluate the enrichment for gene ontology (GO) terms, defined by the GO consortium for functional data mining (Ashburner *et al.* 2000). GO

uses a set of pre-defined key words (ontologies) to describe predictions of gene function and pathways or specific molecules that the gene product might interact with. Such analysis is especially useful for uncharacterized genes since it can potentially uncover specific functions and their overall contribution to a given process, by quantification of the proportion of genes that fall within each GO category. For instance, this approach was used to confirm that somatic, mitotic, meiotic, and postmeiotic expression gene expression clusters of conserved and co-expressed transcripts across species retain genes important for germ-cell differentiation (Chalmel *et al.* 2007a). A different elegant approach was applied by Margolin and co-workers (Margolin *et al.* 2014) that performed gene predictions of unannotated transcripts by comparing RNA-Seq reads to those expected from gene models of the Ensembl and Genscan databases. This approach led to a list of 59 candidates and the identification of a novel gene previously annotated as non-coding. More sophisticated frameworks have since been developed and are quickly evolving towards an integrative analysis of testis biology (Reviewed in Com *et al.* (2014)), in line with what was discussed by Fang and Casadevall (2011) and Payne (2015). Some highlights include the: i) identification of factors involved in the post-meiotic packaging and programming of the male genome and their functional relationship by GO integrated analysis of proteome and transcriptome data (Govin *et al.* 2012); ii) description of five major post-transcriptional regulatory mechanisms governing germ cell development based on an integrative analysis of differential gene and protein expression in different germ cell types (Gan *et al.* 2013); iii) discovery of 44 novel protein-coding loci, with intermediate lncRNA and mRNA genomic features, among a pool of 3559 testicular unannotated transcripts (TUTs) and 506 lncRNAs from rat spermatocytes and spermatids using a proteomics informed by transcriptomics (PIT) strategy (Chocu *et al.* 2014); and iv) novel insights into the Sertoli-germ cell crosstalk, including the validation of two from many identified potential interacting partners located on the surface of Sertoli cells, by studying the germ cell secretome in the rat and sheep with a combination of proteomics, transcriptomics, genomics and interactomics data (Chalmel *et al.* 2014). In parallel, much attention has been drawn into the robustness and accuracy of bioinformatics tools for integrative analysis of functional data from different omics platforms (Reviewed in Haider and Pal (2013); Mooney and McWeeney (2014)).

The promising technique of Ribosome Profiling

Interestingly, efforts at integration of mRNA and protein data revealed a strikingly poor correlation between their abundance levels (Gygi *et al.* 1999; G. Chen *et al.* 2002; Conrads *et*

al. 2005). If we *a priori* exclude experimental bias, any factor affecting mRNA fate (splicing, degradation, transportation and storage), translational efficiency and protein degradation can impact mRNA-protein correlation (Reviewed in Idler and Yan (2012); Vogel and Marcotte (2012); Haider and Pal (2013)). Not surprisingly, the testis stands out as the tissue displaying the weakest correlation between transcriptome and proteome data (Cagney *et al.* 2005). This is in part due to the extensive translational regulation of mRNAs during spermatogenesis [See *Background*; Reviewed in Idler and Yan (2012); de Mateo and Sassone-Corsi (2014); Yadav and Kotaja (2014)], which is also responsible for a temporal discrepancy between transcript production and translation recorded for different germ cell types (Gan *et al.* 2013). One notorious example of the role of temporal translational regulation is the silencing of protamine transcripts. Protamines replace histones in the condensing chromatin of elongating spermatids but are transcribed earlier in meiotic cells and held inactive prior to translation (Kleene *et al.* 1984). Premature translation of these mRNAs leads to abnormal spermatogenesis and infertility (K. Lee *et al.* 1995). In fact, the translational regulatory mechanism by RNA-binding proteins has been shown to control over 700 transcripts in the mouse testis (Iguchi *et al.* 2006). To address this layer of regulation, the authors obtained microarray expression profiles of ribonucleoproteins (RNPs)- or polysome-bound mRNAs from prepubertal testes highly enriched for meiotic (17 dpp) or postmeiotic cells (22 dpp) and compared them with profiles of adult testis. They found that at least 20% of meiotic transcripts shift between RNPs and polysomes, and vice-versa, representing meiotic transcripts that are translationally up or downregulated, respectively. Most displayed an upregulation of translation during late spermiogenesis (RNPs to polysome shift), compensating for the transcriptional silencing from mid-spermiogenesis onwards. Thus, investigating spermatogenesis translatome has the potential to bridge the gap between transcriptome and proteome data but an extensive and comprehensive large-scale investigation of translation regulation in different male germ cells of the adult testis is still lacking.

Traditionally, the dynamics of translation has been studied by polysome profiling (Faye *et al.* 2014; Zuccotti and Modelska 2016). In this technique, cells are treated with a translation inhibitor [e.g. cycloheximide (CHX)] and lysates are passed through a sucrose gradient that separates fractions of free, monosome- and polysome-associated transcripts by ultracentrifugation, which can be further analyzed by different techniques. Measurement of the mRNA distribution patterns obtained provides information about ribosomal occupancy (percentage of transcripts associated with ribosomes) and density (number of ribosomes associated with the mRNA). This allows the estimation of translation efficiency of different

transcripts in normal states or the detection of translational alterations as a response to system perturbations (Zuccotti and Modelska 2016). Although very informative, polysome profiling is an extremely challenging technique that requires skilled investigators to reliably obtain high-quality, intact polysomes from sucrose gradients. Failure to properly separate transcripts with different ribosome loads can compromise the dynamic range of polysome profiling experiments (Ingolia *et al.* 2012).

A new method that allows for deep-sequencing of mRNAs undergoing translation was recently developed (Ingolia *et al.* 2009) and is referred to as Ribosome Profiling. Essentially, translation is halted (by drug treatment or flash freezing) and cell lysates are digested with RNase to generate ribosome-protected mRNA fragments (RPFs) which are then used as template for library preparation of mRNAs undergoing translation. The resulting high-throughput sequencing data is a valuable resource for investigating genome-wide measurements of ribosome occupancy of mRNAs, translation rates, programmed translation of non-canonical isoforms and non-coding RNAs, and translational responses to stress conditions (Ingolia *et al.* 2009; Ingolia *et al.* 2011; Gerashchenko *et al.* 2012; Ingolia *et al.* 2014). Additionally, ribosome profiling of cells treated with different drugs to halt translation during elongation (Cycloheximide) or initiation (Harringtonine and Lactomidomycin) allows the detection and quantification of alternative initiation sites, upstream open reading frames (uORFs) and the kinetics of translation (Ingolia *et al.* 2011; Ingolia *et al.* 2012; S. Lee *et al.* 2012). Other biological questions can then be addressed, such as mRNA transport/translation and translational efficiency of different transcripts as well as of differential translation, by the combined analysis of transcriptome and translatome data from the same biological samples (Ingolia *et al.* 2009; Ingolia *et al.* 2012; Larsson *et al.* 2013). For all the above, ribosome profiling is a very powerful technique and has now been applied to a vast collection of cell types and organisms (Jackson and Standart 2015). Consequently, optimizations of the original protocol (Aeschimann *et al.* 2015; Chung *et al.* 2015; Miettinen and Bjorklund 2015; Reid *et al.* 2015) and methods for data analysis (Zupanic *et al.* 2014; Chung *et al.* 2015; Michel *et al.* 2016; Spealman *et al.* 2016; Xiao *et al.* 2016) have started to emerge and contribute to an overall improvement and simplification of the method. For instance, the use of alternative/additional nucleases allows to capture different features of translation regulation (Miettinen and Bjorklund 2015) and eliminate rRNA contamination without the need for a labor intensive rRNA depletion with biotinylated oligos (Chung *et al.* 2015; Reid *et al.* 2015). Reid and co-workers (Reid *et al.* 2015) described a simplified Ribosome profiling protocol with significant reduction in cost per sample (80%) and preparation time (2 days). Despite the great contribution of these works, a

simplified ribosome profiling protocol that can be applied to reduced cell numbers/low RNA amounts is still lacking. This has been hindering the application of ribosome profiling to testicular germ cells in specific developmental stages. In fact, testicular ribosome profiles have only been generated for whole testis of wild-type (wt) and knock-out (KO) mice (Castaneda *et al.* 2014). Such development would be of the outmost importance for male reproductive biology since it would provide the means to identify all transcripts undergoing translational regulation and investigate at large scales the mechanisms regulating spermatogenesis progression.

Specific Objectives & Results

1- **Isolation of stage-specific male germ cells by Ho-FACS in different species**

Given the high complexity of the testicular tissue, detailed molecular studies of spermatogenesis call for robust, straightforward and reproducible techniques to isolate germ cells in specific developmental stages. Also, the field of male reproductive biology is still lacking a method that could be transversely applied to different species to facilitate analysis in comparative studies. The first part of this chapter describes the efforts to optimize a Ho-FACS based protocol for the isolation of male germ cells from testicular cell suspensions obtained from mouse, rat, dog and frog.

Paper IV - Lima, A.C.*, Jung, M.*, Rusch, J., Usmani, A., Lopes, A.M., Conrad, D.F. *Multispecies purification of testicular germ cells.* Biology of Reproduction (***in revision***)

2- **Development and application of a new ribosome profiling-based technique to Ho-FACS isolated murine male germ cells**

Collectively, both GWAS and candidate-gene approaches for the identification of causal variants of male infertility have been demonstrating the complex genetic architecture of this condition. The low success of such studies is due, in part, to the challenge of attributing functional relevance to disease-associated variants, especially when little is known about the genetic region. Providing functional annotation of genetic variants relies therefore on the knowledge of the basic biology underlying the system under study. In the second part of this chapter I sought to investigate the molecular networks controlling spermatogenesis progression, by identifying the proteins being translated at specific stages. For that, I describe a protocol for ribosome profiling of 4 subtypes of male germ cells, optimized for samples with low RNA input, which captures the dynamics of translational regulation in these cells.

Paper V – Lima, A.C., Yuan, N.H.R., Amorim, A., Conrad, D.F., Lopes, A.M. *Challenges and solutions for ribosome profiling with limited cell numbers: the example of murine male germ cells* (***in preparation***)

Combining these two approaches, my main aim in this chapter was to design and optimize techniques that would allow for deep molecular studies of spermatogenesis progression. This work provides the tools to bring new knowledge in the field of male reproductive biology and to help elucidate the functional impact of genetic variation found in infertile men.

Paper IV - Multispecies purification of testicular germ cells

*Biology of Reproduction (**in revision**)*

Multispecies purification of testicular germ cells

Ana C. Lima[1,2,3,4*], Min Jung[1*], Jannette Rusch[1], Abul Usmani[1], Alexandra Lopes[3,4], Donald F. Conrad[1,5]

[1]Department of Genetics, [5]Department of Pathology & Immunology, Washington University School of Medicine, St. Louis, MO 63110, USA; [2]Graduate Program in Areas of Basic and Applied Biology (GABBA), Abel Salazar Institute of Biomedical Sciences, University of Porto, 4050-313 Porto, Portugal ; [3]Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Porto, Portugal – I3S; [4]Instituto de Patologia e Imunologia Molecular da Universidade do Porto, Porto, Portugal – IPATIMUP, 4200-465 Porto, Portugal

Running Title: Multispecies purification of testicular germ cells

Summary: We investigate whether a popular method for purifying testicular germ cells from mouse will work when applied to other vertebrate species.

Key words: germ cells, FACS, Hoechst 33342, *canis familiaris*, *xenopus laevis*, *rattus norvegicus*, comparative biology, spermatogenesis

Corresponding Author:

Dr. Donald F. Conrad
Department of Genetics
Washington University School of Medicine
Campus Box 8232
St. Louis MO, 63110, USA
E-mail: dconrad@genetics.wustl.edu
Phone: +1-314-362-4379
FAX: +1-314-362-7855

Abstract

Advanced methods of cellular purification are required to apply genome technology to the study of spermatogenesis. One approach, based on flow cytometry of murine testicular cells stained with Hoechst-33342 (Ho-FACS), has been extensively optimized and currently allows the isolation of 9 germ cell types. This staining technique is straightforward to implement, highly effective at purifying specific germ cell types and yields sufficient cell numbers for high throughput studies. Ho-FACS is a technique that does not require species-specific markers, but whose applicability to other species is unknown. We hypothesized that, due to the similar cell physiology of spermatogenesis across mammals, Ho-FACS could be used to produce highly purified subpopulations of germ cells in mammals other than mouse. To test this hypothesis, we applied Ho-FACS to two mammalian species that are widely used in testis research (*Rattus norvegicus* and *Canis familiaris*) and, selected, as a vertebrate outgroup, a polyploid model organism (allotetraploid *Xenopus laevis*). We successfully isolated 4 male germ cell populations from dog and rat testes with average purity ~80% estimated by microscopy. We were unable to purify distinct germ cell populations from frog and discuss the limitations of adapting Ho-FACS to this non-mammalian species. Additionally, we propose an optimized gating strategy to better discriminate round and elongating spermatids in the mouse, which can potentially be applied to other species. Our work indicates that spermatogenesis may be uniquely accessible among mammalian developmental systems, as a single set of reagents may be sufficient to study over 5,000 mammalian species, opening a new avenue in the fields of development and male reproductive biology.

Introduction

Spermatogenesis is a complex developmental process in which early spermatogonial stem cells differentiate into spermatozoa in the seminiferous tubules of the testes. The study of this fascinating process has produced critical insights into stem cell biology (Chen *et al.* 2005), developmental gene regulation (Soumillon *et al.* 2013), adaptive evolution (Carelli *et al.* 2016) and fertility (Good *et al.* 2010). With over 30 different distinct cell types in the vertebrate testis, there is exceptional diversity in the expression profiles of cells within a single individual, which can become confounding when studying expression differences among individuals or developmental stages (Rodriguez-Casuriaga *et al.* 2013). This has compelled researchers to develop methods for effective male germ cell enrichment and isolation, such as StaPut velocity sedimentation, elutriation, magnetic-activated cell sorting (MACS), whole testis collection during first wave of spermatogenesis and fluorescence-activated cell sorting (FACS) with Hoechst-33342 (Hoechst).

StaPut and elutriation are fairly efficient techniques that allow separation of different germ cells based on their size and density. When applied to mouse, StaPut yields about 10^8 cells/population from 22 testes with 90% purity, whereas approximately 10^7 cells/population can be obtained by elutriation of two testes with 80-95% purity rate (Chang *et al.* 2011; Bryant *et al.* 2013). In both methods, the fractionation step that collects purified cells from different BSA or Percoll gradients is labor intensive (3-4 hours) and requires proficiency from practice as well as specific equipment. Also, both techniques are unsuitable for detailed molecular studies during meiosis as they can only separate one type of meiotic cell subpopulation at a time and fail to yield sufficient purity (Namekawa *et al.* 2006; Getun *et al.* 2011). MACS, which separates desired germ cell populations by conjugating the germ cells with a known surface marker antibody primed with magnetic beads, may circumvent this issue by performing purification in parallel with population-specific antibodies. However, only spermatogonia and spermatids are proven to have established surface markers for successful enrichment (Bryant *et al.* 2013). Furthermore, antibody-assisted purification has limitations in that it be necessary to develop species-specific reagents for each marker, and antibody-assisted purification typically does have the sensitivity to discriminate between cells at slightly different stages of a quantitative developmental process. Collecting mouse testis samples at specific days post-partum (dpp), timed for the first appearance of different germ cell-types during first wave of spermatogenesis, is also used to enrich specific germ cell populations (Yoshida *et al.* 2006). Given that the testis size is very small at those time points, and that samples comprise a mixture

of all testicular cells, this approach is experimentally challenging and fails to detect intrinsic biological variations among individual cells. Importantly, evidence from different studies suggests that the first wave is regulated differently from adult spermatogenesis (Yoshida *et al.* 2006; Laiho *et al.* 2013; Margolin *et al.* 2014).

FACS of Hoechst stained (Ho-FACS) murine male germ cells can discriminate 9 germ cell-types (Lassalle *et al.* 2004; Bastos *et al.* 2005; Getun *et al.* 2011; Gaysinskaya *et al.* 2014; Gaysinskaya and Bortvin 2015). Hoechst-33342 is a vital dye that binds preferentially to poly(d[AT]) sequences in the minor groove of DNA, with secondary binding taking place at higher ratios. These two DNA binding sites show varying binding energies and consequent spectrum shifts in relation to chromatin amount and structure (Sandhu *et al.* 1985; Watson *et al.* 1985). It has been proposed that this spectral shift could be used to discriminate between cells with similar DNA content but different chromatin properties (Smith *et al.* 1985; Steen and Stokke 1986; Ellwart and Dormer 1990). Indeed, Ho-FACS of male germ cells exhibits a pattern that reflects changes in DNA content (blue fluorescence) and chromatin structure (red fluorescence) throughout spermatogenesis. In fact, red fluorescence shifts resulting from progressive chromatin de-condensation during meiotic prophase allow the resolution of different meiotic subpopulations (Bastos *et al.* 2005; Gaysinskaya *et al.* 2014). Spermatogonial stem cells are an exception and represent a side population due to BCRP1-dependent dye efflux, which is switched off after the spermatogonial stages (Bastos *et al.* 2005). With over 95% purity of sorted populations (Gaysinskaya *et al.* 2014) and an average of 10^7 cells/ population from two testes in less than two hours, this technique has proven highly efficient and less labor intensive. Although the actual FACS session requires specialized sorting equipment (UV laser) and a skilled operator, many research facilities provide cell-sorting services. Currently, Ho-FACS is more favorable for high-throughput studies in spermatogenesis in comparison to other techniques as it provides higher yield and purity of the isolated germ cell-types.

Recently, there is a growing interest in applying genomic technology to germ cells, especially in an evolutionary context (Liu *et al.* 2015; McCarrey 2015). In that sense, purified cells can be used for numerous applications ranging from studying gene regulation, nucleosome mapping, epigenetics, development in germ cells and many more (Chowdhury *et al.* 2009; Getun *et al.* 2010; Roig *et al.* 2010). To unravel the complexity of germ cells at a genomic level, researchers need an efficient and high-throughput purification technique that can be applied easily, to many species. Given that Ho-FACS is not based on a species-specific molecular signature (e.g. an

antigen), and that the cellular machinery for spermatogenesis is similar across all vertebrates, we hypothesized that separation of different germ cell types by Ho-FACS could be applied to other species. To test this hypothesis, we applied Ho-FACS to two species that are highly valued by the testis research community: *Rattus norvegicus* (rat) and *Canis familiaris* (dog). Finally, we wished to explore whether Ho-FACS might easily allow germ cell purification from all vertebrates. A general feature of vertebrate spermatogenesis is the synchronized and stepwise differentiation of male germ cells; thus, we hypothesized that applying Ho-FACS to any vertebrate testis would produce the type of structured cytogram necessary to recognize and purify distinct germ cell stages. Thus, we sorted testicular cells from an “outgroup” vertebrate model organism, *Xenopus laevis*, which is quite distinct from mammals as it has a polyploid genome and cystic (instead of tubular) testis.

Our results provide detailed descriptions on how Ho-FACS performs at cell enrichment for 4 primary types of germ cells in each of the 3 species that we investigate. While the tetraploid frog sample lacked clearly defined cell populations, each of our target spermatogenic germ cell types could be distinguished by Ho-FACS of the diploid mammalian species. We also demonstrate the use of species-specific conditions for testis dissociation, and present an optimized FACS gating strategy based on cell shape, size and complexity to distinguish elongated and round spermatids in mouse. Collectively, we offer the first proof of principle that flow cytometry can be applied transversally across mammalian species to isolate Hoechst stained male germ cells in different developmental stages.

Materials and Methods

Animals

C57BL/6 male mice (Jackson Laboratory) and Sprague Dawley male rats (Harlan Bioproducts) were raised in animal facilities at Washington University in St. Louis. Male frog (*Xenopus laevis*, Nasco, #LM00715M) testes were obtained at Dr. Zhongsheng You (Washington University in St. Louis). Dog testes (Labrador-Pitbull and Terrier mixed breeds) were collected at Hillside Animal Hospital (St. Louis, MO), from animals scheduled to perform castration, and were transported to the lab on ice for immediate processing. Prior to surgery, dogs are routinely injected with lidocaine and bupivacaine to help with the recovery process. All testis samples were obtained from sexually mature animals (Mice: 8 - 12 weeks, Rats: 70 days, Dogs: 12-24 months, Frog: 10-12months) and procedures were conducted in compliance with regulations

of the Animal Studies Committee at Washington University in St. Louis

Collection and processing of testicular tissue

Fresh testes from each species were de-capsulated, rinsed in 1X Phosphate buffered saline (PBS; Thermo Scientific #AM9625) and cut to the size of mouse testis (approximately 1.5cm X 0.7cm). Sections were used without further processing for enzymatic dissociation and FACS sorting or fixed for histology. For immunofluorescence, tissue was fixed in 4% Paraformaldehyde (PFA; VWR #15710) overnight at 4°C and washed with 70% EtOH at least 3 times. Testes sections used for Hematoxylin-Eosin (HE) staining were collected in Modified Davidson's solutions (24h at room temperature with gentle rotation; Electron Microscopy Sciences #64133-50), fixed in Bouin's solution (24h at room temperature with gentle rotation; Sigma #HT101128) and washed with 70% EtOH until any remaining yellow color of Bouin's fixative was completely removed.

Immunofluorescence and Hematoxylin-Eosin Staining

Fixed testes samples were processed in an ethanol series, embedded in paraffin and 5µm sections were cut. Slides were de-paraffinized with xylene and rehydrated to PBS through sequential ethanol washes with decreasing alcohol concentrations. Standard HE staining was performed according to HE protocol adapted from Belinda Dana (Department of Ophthalmology, Washington University in St. Louis School of Medicine) with Hematoxylin 560 (Surgipath #3801570) and 1% Alcoholic Eosin Y 515 (Surgipath #3801615) for overall morphological evaluations. Immunofluorescence staining was performed after antigen retrieval (boiling in citric acid buffer for 20min) and tissue permeabilization/blocking (0.5% triton x-100 + 2% goat serum in 1X PBS for 1h at room temperature). Primary [anti-P-H3(ser10); AbCam, #Ab5176] and secondary (goat anti-rabbit ALF 633; Life Technologies #A21071) antibodies were diluted (1:100; 1:500 respectively) in antibody dilution buffer (1X PBS + 1% Tween 20 + 1% BSA) and incubated overnight at 4°C and 4h at room temperature, respectively, in a humid chamber. After secondary antibody incubation, sections were stained with Hoechst-33342 (1:500; Life Technologies, #H3570), washed with 1X PBS and mounted with ProLong Diamond Antifade Mountant (Life Technologies #P36961).

Testis dissociation and Hoechst Staining

A multi-species testis dissociation protocol was designed based on the procedure described in Getun et al. 2011 for mouse.

Preparation of solutions (fresh; prior to testes collections): Collagenase type I (120U/mL; Worthington Biochemical, #LS004196) + Cycloheximide (0.1mg/mL; Amresco #94217) in 1X Dulbecco's Modified Eagle Medium (DMEM; Life Technologies #31053). Trypsin (50mg/mL; Worthington #LS003708) in 10mM HCl. DNase I (1mg/mL; Roche #10104159001) in 50% glycerol.

1- *Testis enzymatic digestion*: Testes (mouse) or testes sections (rat, dog, and frog) were placed in 15mL conical tubes containing 3mL of DMEM/Collagenase I/CHX solution, and 10µL of DNase I solution. The tube was shaken vigorously until the testicular tubules started to disperse and then agitated horizontally at a speed of 120 for 15 min at 33°C. Temperature and agitation speed were the same for all subsequent incubation steps.
2- *Somatic cell removal*: Tubules were decanted for 1 min vertically at room temperature and the supernatant was discarded to remove somatic cells.
3- *Seminiferous tubule digestion*: 2.0 mL of DMEM/Collagenase I/CHX, 50µL of Trypsin and 10µl of DNase I solutions were added and the tube was inverted several times. After 15 min incubation period, the tubules were gently pipetted up and down for 3 min using a plastic disposable Pasteur pipet with wide orifice. Then, 30µL of trypsin and 10µL of DNase I solution were added and the tube was inverted several times, followed by another 15 min digestion period.
4- *Staining with Hoechst*: 400µL of Fetal Bovine Serum (FBS; Thermo Scientific #10082139) was added and mixed by inverting to inactive trypsin, followed by addition of 5µL of Hoechst-33342 (Hoechst; Life Technologies, #H3570) and 10 µL of DNase I. The cell suspension was incubated for 15 min, then passed through two 40um 1X DMEM pre-wetted disposable filters and stored on ice and protected from light until ready for FACS processing (not more than 30min).

In order to evaluate the effect of testis dissociation protocols in FACS, tissue sections of all specimens were also dissociated using species-specific protocols based on the procedure described above with the following modifications:

Rat: Trypsin stock concentration was adjusted to 1mg/ml and Hyaluronidase (1mg/ml in 1X DMEM; Sigma-Aldrich #H6254) stock solution was added to the DMEM/Collagenase I/CHX solution. The last incubation time in step 4 was adjusted to 20 min.

Dog: Collagenase stock concentration was adjusted to 0.2% by dissolving 20mg Collagenase type I in 10ml 1X DMEM. The first incubation time in seminiferous tubule digestion was adjusted to 30min.

Frog: Hyaluronidase (0.5 mg/ml) was added to the DMEM/Collagenase I/CHX solution. All incubations were carried out at 27°C and the incubation time in step 4 was adjusted to 25 min. After seminiferous tubule digestion, we repeated triturating the suspension with a disposable Pasteur pipette, and adding trypsin/DNAseI enzyme mix. Then the suspension was incubated for 15 min at 27°C at 120rpm.

Fluorescence Activated Cell Sorting (FACS)

Cells were sorted and analyzed by a Beckman Coulter MoFlo Legacy cell sorter, using Summit Cell Sorting software, similarly to the descriptions of (Getun *et al.* 2011). Hoechst was excited using a ultra-violet laser and triggered for scatter by a 488nm blue laser. To detect Hoechst's wide emission spectrum, the U.V. laser was paired with a 463/25nm band pass filter for detecting Hoechst Blue and a 680nm LP band pass filter for Hoechst Red. A 555DLP dichroic was also used to distinguish blue from red emission wavelengths. Samples were analyzed using a 70-micron nozzle and the sorting flow rate was set to 3,000-4,000 events/second. A minimum of 500,000 events were detected before proceeding to gating. Since we did not use Propidium Iodide to exclude non-viable cells, we engaged in a sequential cell gating strategy: debris was excluded based on FSC vs SSC plot, then singlets gated by adjusting threshold for forward scatter pulse width and finally red/blue Hoechst fluorescence was used to detect different spermatogenic germ cell populations. Each testis was processed for 1.5 hours to collect an average of 6.0 X 10^6 cells for each subpopulation. Cells were collected into 1 mL of 1X PBS + 10% FBS in 5 mL polypropylene round-bottom tubes that were pre-coated with FBS. To concentrate the samples and remove dead cells and cellular debris, sorted cells were pelleted by centrifugation (600xg at 4°C for 10min) and washed in 1mL of ice-cold 1X PBS.

Microscopic evaluation of purified cells

To identify the cell types gated in each FACS sorted population we evaluated chromatin structure and cell morphology microscopically, based on Hoechst fluorescence. During the wash step after FACS, 100 µL of sorted cells were collected, fixed in 4% PFA and stored at 4°C in the dark. Slides were mounted with 20 µL of fixed cells from each population and visualized in upright confocal or light microscopes. To quantify cell purity, images were obtained from a minimum of 5-15 random fields and/or at least 100 cells (when available) were counted to estimate contamination with other cell types. To avoid human errors, cells were counted independently by two researchers and an estimated average cell count was reported.

Results

Efficiency of tissue dissociation protocol is crucial for cell sorting with Hoechst staining

We isolated two testes from each animal in the study: 10 mice, 2 rats, 2 dogs, 1 frog. In order to confirm a normal adult testis phenotype of the collected specimens, we performed Hematoxylin and Eosin (HE) staining of tissue sections from one testis, and then submitted the other testis for FACS. A microscopy analysis of the HE slides shows the expected tissue architecture and organization of normal adult male testes and highlights some general differences across species (Fig. 1). In the frog, germ cells are organized in cysts, with each cluster containing cells in the same developmental stage. Mammals have a tubular testicular arrangement, with spermatogenesis progressing from the periphery towards the lumen, and show interspecific variability of germ cell morphology.

The success of cell sorting protocols depends on the quality of inputted single-cell suspensions, and is therefore directly affected by the efficiency of tissue dissociation. Here, we evaluated the use of different protocols for testis enzymatic dissociation in each of the species studied. We applied an enzymatic dissociation protocol optimized for mouse testis to all species, referred to as the “multispecies” protocol, and defined species-specific protocols by adjusting incubation temperatures and times, trypsin concentrations and/or introducing the use of hyaluronidase to improve digestion of connective tissues (Methods). To control for technical and biological variables, the experiments were performed simultaneously in tissue sections of the same testis

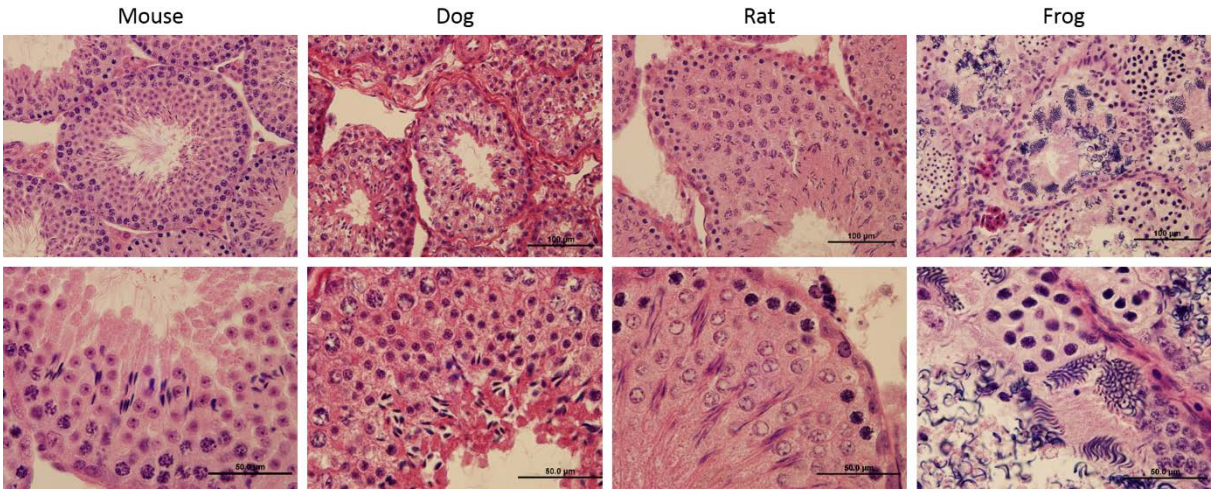


Figure 1. Hematoxylin and Eosin staining of testicular cross-sections of collected specimens.
For each animal studied in the paper (n=10 mice, n=2 dogs, n=2 rats, n=1 frog) we processed one testis for histology and one for FACS. Here we present representative H&E staining from each species. In each subject, histological examination of testicular cross-sections shows the presence of all germ cell-types in different developmental stages, confirming that the specimens were sexually mature and presented a normal testicular phenotype. Cellular arrangement of the frog testis is strikingly different from mammalian species, with germ cells developing in cysts instead of linearly progressing in the direction of the tubule lumen. Lower panel represents higher magnification (100X) of images displayed in upper panel (40X).

for both biological replicates, except for frog. Overall, species-specific dissociation protocols performed better as evaluated by the separation of distinct clusters obtained on the FACS profiles. In fact, for the frog testis we observed cellular aggregates with the multispecies protocol, an indication of incomplete dissociation, making the sample unusable for flow cytometry. FACS plots of dog and rat obtained using the multispecies protocol dissociation do not exhibit the typical Ho-FACS profile and are marked by the absence of some germ cell subpopulations (Fig. S1). To evaluate the performance of our approach, we estimated the mean percentage of cells that passed the gates during FACS in these two species and compared these values with the ones obtained for mouse. The proportion of live cells from total counts (Fig. 2A) is indicative of the sample quality and a reflection of the efficiency of tissue dissociation protocols. The average percentage of cells passing the debris filter (live cells) was comparable between species with the exception of frog, for which a low number of live cells (21%) suggests that the dissociation protocol requires further improvements. Notably, the percentages estimated for all species are directly influenced by the high stringency of the debris filter applied during FACS. For all species, most live cells were singlets (Fig. 2B). These results suggest that our adjusted protocols are effective in generating good single-cell suspensions, but might

require small adjustments to keep cellular integrity throughout the dissociation process. Also, cell sorting with Hoechst staining seems very sensitive to sample quality, validating our approach of designing species-specific dissociation protocols.

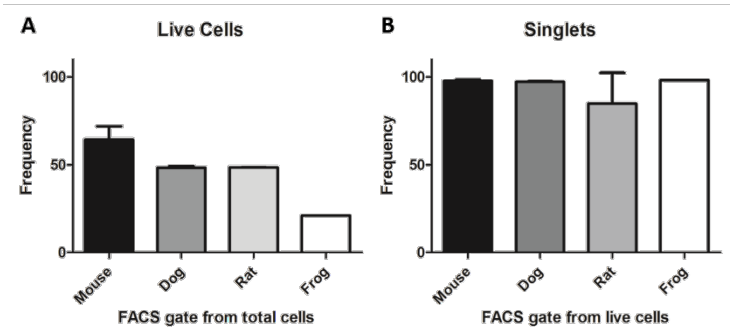


Figure 2. Evaluation of species-specific testis dissociation protocols by flow cytometry.
In order to evaluate the efficiency of dissociation protocols in different species we estimated the percentage of live cells from total number of cells (A) and the proportion of those that were single-cells (B). The quality of dog and rat single-cell suspensions was similar and slightly inferior to the average mouse sample (A). The low number of live cells recovered for the frog (21%) suggests that the dissociation protocol requires further improvements. Almost all live cells are singlets, with the lower average detected for rat (B). Datasets were obtained from variable numbers of experimental replicates (mouse: 10; dog and rat: 2 and frog: 1) using FlowJo® software v10 (Tree Star Inc.). Histograms were generated with GraphPad Prism (version 5.02 for Windows, GraphPad Software, San Diego California USA, www.graphpad.com), plotting the calculated mean values with standard deviation.

Male germ cell-types of different mammalian species can be discriminated by Ho-FACS

Figure 3 shows the Ho-FACS cytograms obtained for dog and rat male germ cells, in comparison with a typical mouse cytogram. Although we see some expected interspecific variation in the pattern of the FACS profiles, we can clearly distinguish four subpopulations of male germ cells for both species. Thus, we defined the gates for sorting based on the cluster of cells observed and taking into account the expected location of the subpopulations in terms of Ho red and blue fluorescence: i) spermatogonia (Spg; side population)- low Hoechst blue and red fluorescence; ii) primary spermatocytes (Spc I; 4N eu- to heterochromatin) – high Hoechst blue and a wide range of low to high Ho red fluorescence; iii) secondary spermatocytes (Spc II; 2N eu- to heterochromatin)- intermediate Hoechst blue and a wide range of low to high Ho red fluorescence; iv) spermatids (Spd; 1N compacted chromatin with structural variations resulting from histone to protamine transition) – low Hoechst blue and a narrow range of Ho

red. Moreover, it appears that the chromatin of the dog germ cells is generally more compact throughout spermatogenesis as the clusters of cells show a trend of lower red Hoechst fluorescence.

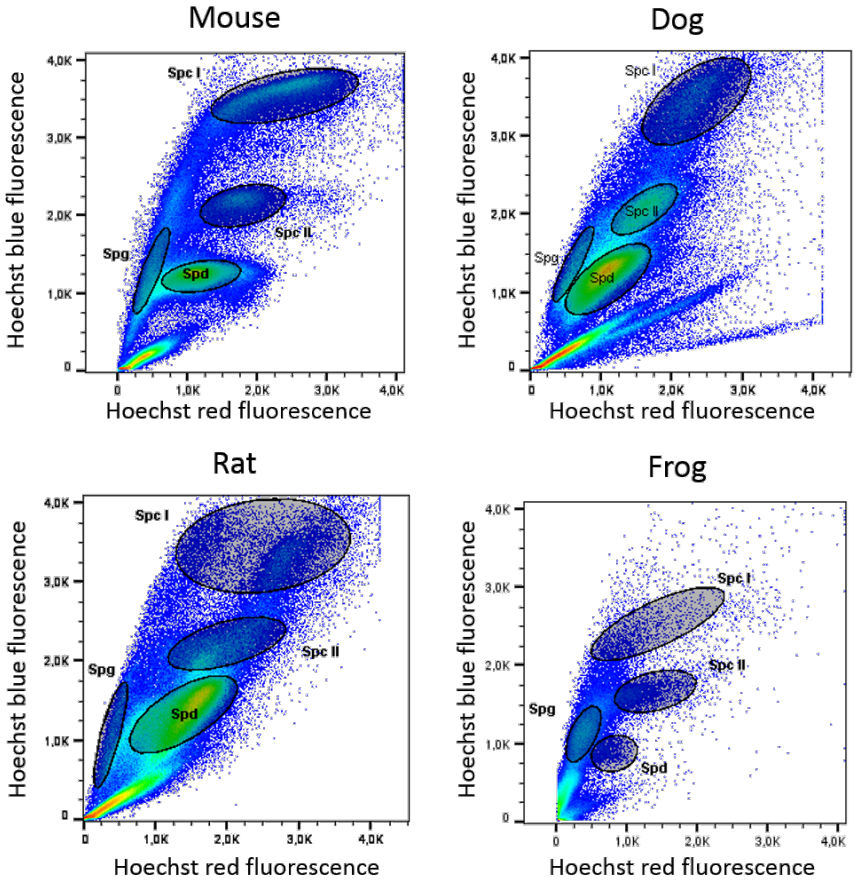


Figure 3. Interspecific comparison of Ho-FACS plots of testicular single-cell suspensions.

The binding of Hoechst to DNA results in different FACS patterns depending on chromatin compaction and quantity. Plots represent the ratio of blue and red Ho fluorescence obtained by flow cytometry after testis dissociation and staining of germ cells of the mouse, dog, rat and frog. Gating (round circles) was defined based on observed cell clusters and expected location of populations in relation to Ho fluorescence. Despite small interspecific variations, 4 populations were identified and sorted for all diploid species. For the allotetraploid frog only one population is clearly defined (labeled Spg) but the same gating principles were applied. Spg: spermatogonia; Spc I: primary spermatocytes; Spc II: secondary spermatocytes; Spd: spermatids. Plots were obtained using FlowJo® software v10 (Tree Star Inc.).

To identify the germ cell-types and quantify the purity of the sorted subpopulations, we performed a microscopy evaluation of cell morphology and chromatin structure based on

Hoechst fluorescence (Methods). For the frog sample, the lack of clearly defined populations observed during FACS was confirmed microscopically. All sorted populations showed a mixture of cells in different stages (Fig. S2). Immunofluorescence in tissue sections with Hoechst and a marker for mitotic cells, Histone 3 serine 10 phosporylation [P-H3 (ser10)], was used as reference for the pattern of Hoechst staining in different germ cells of dog and rat testes (Fig. 4A). Spg are small, round shaped cells with distinct pericentric heterochromatin. Spermatocytes are larger granulated cells with Spc II populations defined by the detection of binucleated cells or cells in diakinesis. Spd are small haploid cells that can be round or elongated in shape. Despite the similar size, round spermatids can be clearly distinguished from Spg by the presence of localized chromocenters. The images obtained for each population in both dog and rat samples (Fig. 4B) are in accordance with the cell morphology and nuclear structure described above. Purity was estimated based on this analysis, indicating 74%-85% purity of specific cell-types passing through each gate (Fig. 4B). The one exception to this was a low 46% purity of the dog Spg population, due to the close proximity of the eSpd and Spg populations in fluorescence space (Fig. 3 and Fig. 4B). For the Spc I gates, most contamination was with preLeptotene spermatocytes. In the rat, this could have resulted from the absence of clearly defined pre-meiotic and meiotic spermatocytes subpopulations during FACS. Looking at the relative frequency of cells passing through each gate (Fig. 5) a similar higher frequency of spermatids was detected for all species, however, the abundance of other germ cell types varies amongst species. These observations were expected and presumably reflect interspecific differences in testis composition and the technical challenge of making standardized settings for subpopulation gating. Altogether, our results indicate that Hoechst FACS of testis single-cell suspensions can be used to isolate germ cells from dog and rat, further strengthening our hypothesis that this method can potentially be applied as a generalized procedure for isolation of germ cells in all mammalian species.

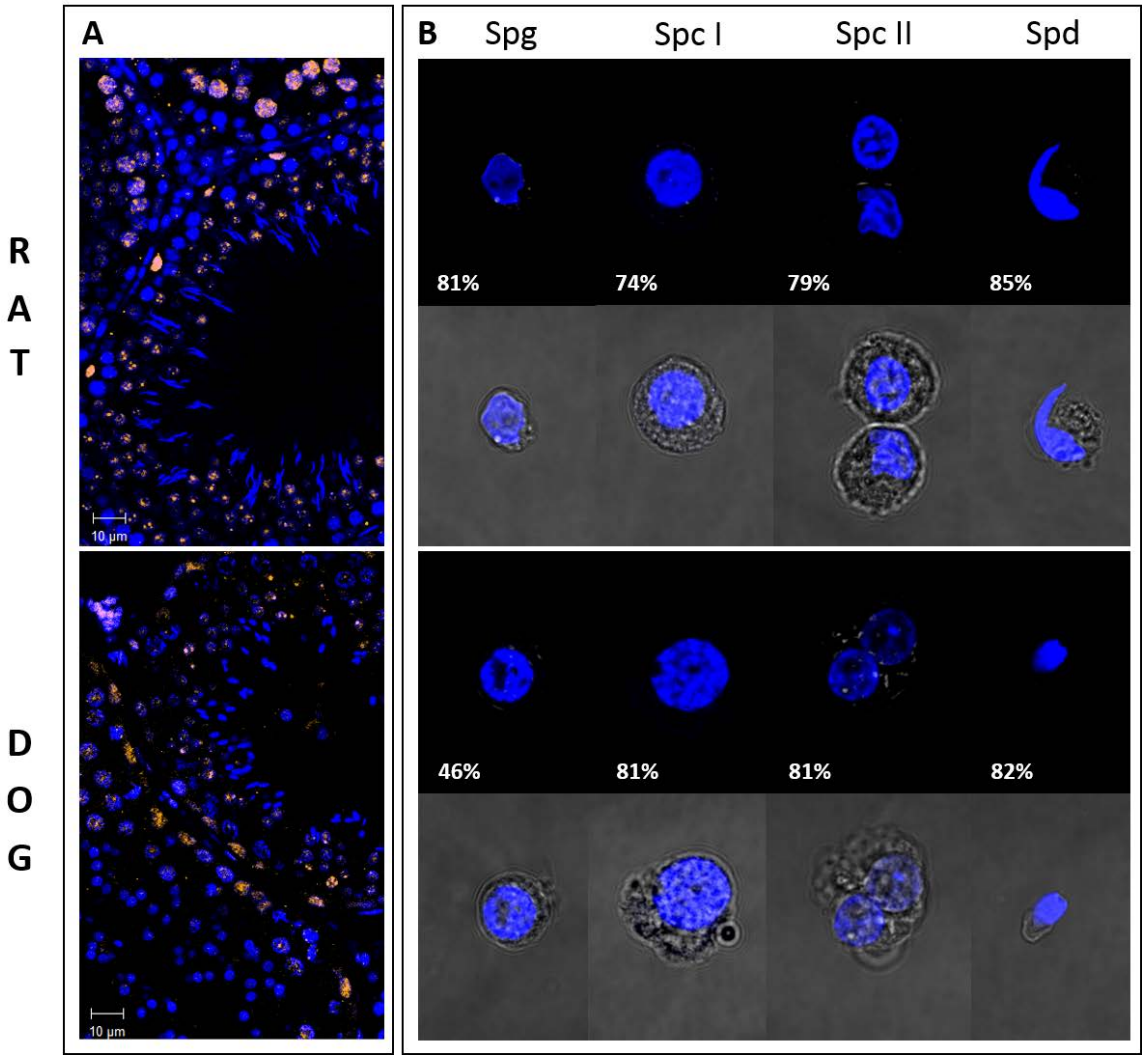


Figure 4. Microscopic evaluation of germ cell populations isolated from dog and rat testis by Ho-FACS. Immunostained cross-sections of dog and rat testes (A) were used as reference for the classification of isolated germ cells in respect to chromatin structure marked with Hoechst (blue). Histone H3 phosporylation (orange) marks mitotic cells. Morphological evaluation of chromatin structure and cell shape and size confirms the enrichment of expected cell-types for each sorted population in both species (B). Small round cells with compact heterochromatin were identified in the Spg gate, indicating the enrichment of spermatogonia in that population. Panels Spc I and Spc II show larger and more complex cells with more diffuse chromatin (primary spermatocytes) and/or binucleated cells (secondary spermatocytes). Spd gate comprised cells in different states of spermiogenesis, ranging from round to completely elongated spermatids. Percentages indicate the estimated purity of each sorted population. Spg: spermatogonia; Spc I: primary spermatocytes; Spc II: secondary spermatocytes; and Spd: spermatids panels indicate the designated FACS gate. Slides were visualized in a confocal microscope. For each isolated population, Hoechst fluorescence of sorted cells was visualized after FACS and images were collected under a 63X magnification lens, with (lower panel) or without (upper panel) white light transmission.

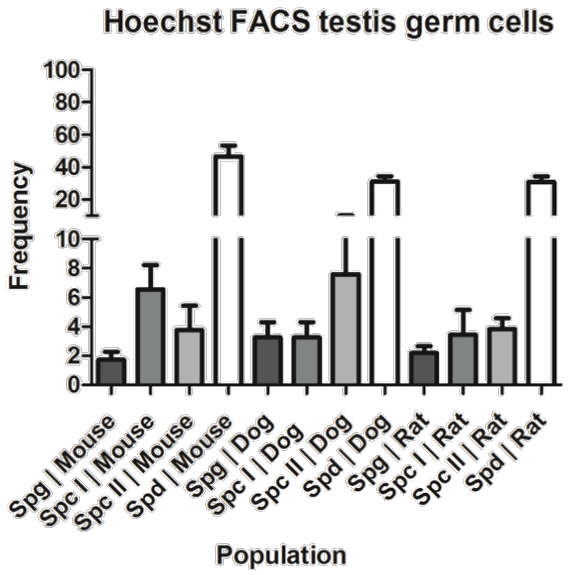


Figure 5. Interspecies comparison of germ cell heterogeneity measured by Ho-FACS.

The cycle of the seminiferous tubule epithelium influences the proportion of germ cell-types populating the tissue. Plotting the percentage of live cells falling in each gated population (C) indicates a high proportion of spermatids for all species and a variable number of other cell-types, reflecting the interspecific dynamics of cell heterogeneity during spermatogenesis. Gated populations were labeled during FACS in respect to the expected cell-type: Spg- spermatogonia; Spc I- primary spermatocytes; Spc II- secondary spermatocytes and Spd- spermatids. Datasets were obtained from variable numbers of experimental replicates (mouse: 10; dog and rat: 2 and frog: 1) using FlowJo® software v10 (Tree Star Inc.) and histograms were generated with GraphPad Prism (version 5.02 for Windows, GraphPad Software, San Diego California USA, www.graphpad.com), plotting the calculated mean values with standard deviation.

Round and elongating spermatids can be separated by Ho-FACS based on cell shape and size

Given that round (rSpd) and elongating (eSpd) spermatids are molecularly very distinct in terms of transcription activity as well as the differentiation occurring in the latter during spermiogenesis, we sought to isolate different mouse spermatid subpopulations by FACS. Hoechst red and blue fluorescence is insufficient to discriminate spermatid subpopulations. However, it has been previously suggested that rSpd and eSpd could be gated based on high forward scatter (FSC high) and low forward scatter (FSC low), respectively (Bastos *et al.* 2005). Interestingly, we observed that gating based on the FSC parameter alone introduced some contamination in the sorted populations. Microscopy quantification of purity of sorted

populations based on cell morphology and Ho fluorescence revealed and enrichment of ~62% for eSpd and 84% for rSpd (Fig. S3). Gating for events with low FSC and high VS low side scatter (SSC), we increased purity levels to 92% of eSpd and 86% rSpd (Fig. S3). Finally, we observed that the lowest levels of contamination could be obtained by the combination of FSC and SSC gating followed by Hoechst red/blue fluorescence. It seems that eSpd can be isolated gating for low FSC&SSC with 83-92% enrichment range, whilst rSpd appear to have higher FSC&SSC values and can be separated with 86-95% accuracy (Fig. 6). Importantly, this gating strategy is based on cell size, shape and complexity and thus potentially applicable to Ho-FACS of any species undergoing spermiogenesis during gamete development.

Discussion

One of the major challenges in male reproductive biology has been to design a method to differentiate and isolate subtypes of developing germ cells with a high percentage of recovery and low contamination with other cell-types. Since the first reports over a decade ago, flow cytometry of Hoechst stained murine male germ cells has been slowly revisited and optimized to isolate pre-meiotic (spermatogonia), meiotic (preleptotene, leptotene, zygotene, pachytene, diplotene and secondary spermatocytes) and post-meiotic (round and elongating spermatids) stages (Lassalle *et al.* 2004; Bastos *et al.* 2005; Shimizu *et al.* 2006; Getun *et al.* 2011; Gaysinskaya *et al.* 2014; Gaysinskaya and Bortvin 2015). We reasoned that, for species sharing similar chromatin dynamics (2N- 4N- 2N-1N) and structure throughout spermatogenesis, major populations of germ cells in different stages could be isolated by Ho-FACS. We have shown here that the Ho-FACS protocol for germ cell purification can be applied broadly to mammalian species beyond mouse.

While for all diploid mammalian species we could detect defined germ cell clusters during Ho-FACS, the African clawed frog sort was notably marked by their absence. This could result from allotetraploidy generating intrinsic differences in chromatin structure, or a fundamental difference in cell physiology between mammals and amphibian germ cells. We also cannot exclude the possibility that an improved dissociation protocol for frog would allow us to observe clearer clusters, as only 20% of the cells recovered from the frog sample were alive (Fig. 2). Nevertheless, in combination, these observations support our hypothesis that the current design of Ho-FACS protocol can successfully be performed in species sharing a similar 2N-

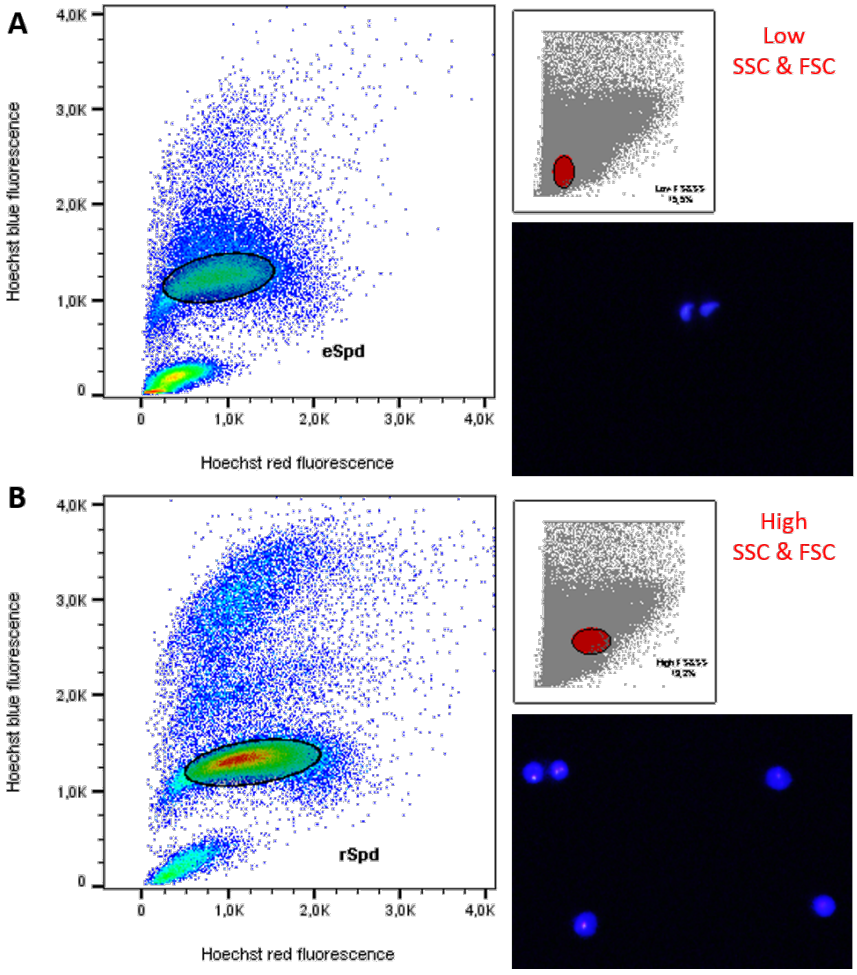


Figure 6. Gating strategy to discriminate round and elongating spermatids.

Cell shape and complexity influence the ratio of forward (FSC) and side scatter (SSC) parameters measured by flow cytometry. The smaller windows in both images show the parent gate (red full circle) based on FSC and SSC. Gated cells then clustered as functions of Hoechst blue/red fluorescence with the pattern expected for haploid cells with condensed chromatin, defining the gates for sorting. Morphology of sorted cells was evaluated microscopically based on Hoechst fluorescence and confirms the enrichment of the expected cell-types in each population. Therefore, elongating spermatids are smaller and less complex showing lower ratios of FSC and SSC (A), whereas round spermatids present higher FSC and SSC (B). Cell images were obtained by light microscopy with a UV lamp (16X magnification lens). Plots were generated using FlowJo® software v10 (Tree Star Inc.).

4N-2N-1N chromatin dynamics. In theory, the applicability of this technique to organisms of different ploidy should be possible but needs to be further investigated.

From the two diploid species, we show that the general resolution of distinct cell populations is maintained across mammals and allows the isolation of at least four developmental stages: spermatogonia, primary spermatocytes, secondary spermatocytes and spermatids. The purity of these subpopulations was slightly reduced when compared to previous works (Bastos *et al.* 2005; Gaysinskaya *et al.* 2014) but shows good enrichment of expected cell-types (Fig.4). The presence of eSpd was the major source of contamination within the Spg gates and resulted from their close special proximity in Hoechst plots, reaching the highest values in the dog FACS. One possible way to circumvent this issue would be to stain germ cells with a membrane permeable marker for the acrosome, allowing to gate cells for the presence of this spermatid-specific structure. Alternatively, an easier method would be to optimize a DNA content exclusion gate to discriminate cells first based on peaks of DNA amount and then on Hoechst blue/red fluorescence patterns (Gaysinskaya *et al.* 2014). This gating strategy could also resolve preleptotene (S-phase) from other spermatocytes, simultaneously reducing the sources of contamination observed in Spc I gates, especially for the rat.

Future work would also include the optimization of this protocol to discriminate other cell-types in different mammals. Here, we describe an optimized gating strategy based on cell size, shape and complexity to differentiate rSpd and eSpd in the mouse (Fig. 6 and Fig. S3), suggesting that the isolation of populations enriched for these germ cells can be achieved for other mammalian species. Also, discrimination between different meiotic stages, already resolved for mouse (Bastos *et al.* 2005; Gaysinskaya *et al.* 2014; Gaysinskaya and Bortvin 2015), would broaden the scope of the application of this technique in the field of male reproductive biology.

Overall, we provide the first evidence supporting the applicability of Ho-FACS as a transversal method to isolate male germ cells in different developmental stages across mammalian species. As a proof of principle, our work has major implications for several types of studies in developmental biology. First, it provides the tools to investigate the dynamics of germ cell development in different species individually, which would benefit research of understudied mammalian species such as domesticated animals (Gonzalez and Dobrinski 2015). Furthermore, using the same experimental procedure in different species reduces noise and eliminates sources of variables which often challenge comparative studies. In the “omics” era, with the growing interest in applying genome technology to address questions about epigenetics, regulation and protein diversity throughout spermatogenesis (Chowdhury *et al.* 2009; Getun *et al.* 2010; Roig *et al.* 2010; Gan *et al.* 2013; Laiho *et al.* 2013; Castaneda *et al.* 2014; Margolin *et al.* 2014; Rathke *et al.* 2014; Schlatt and Ehmcke 2014; Liu *et al.* 2015;

McCarrey 2015), this technology could be used to comprehensively tackle different aspects of germ cell development with an evolutionary perspective.

Acknowledgements

We thank Dr. Zhongsheng You at Washington University in St. Louis for frog testes; Hillside Animal Hospital, (St. Louis, MO) for dog testes; Jason Arand and Dr. Ted Cicero's lab at Washington University in St. Louis for providing rat testes and Brianne Tabers for helping with the collection. We thank the Alvin J. Siteman Cancer Center at Washington University School of Medicine and Barnes-Jewish Hospital in St. Louis, Mo., for the use of the Siteman Flow Cytometry Core, which provided staff-operated cell sorting service. The Siteman Cancer Center is supported in part by an NCI Cancer Center Support Grant #P30 CA91842.

Funding

This research was funded by an FCT doctoral fellowship [SFRH/BD/51695/2011 to ACL], grants from the United States National Institutes of Health [R01HD078641 and R01MH101810 to DFC] and an FCT research contract [IF/01262/2014 to AML]

Competing Interests

All authors declare no competing interests.

Author's contributions

ACL and MJ obtained and processed tissue. ACL, MJ, AU and JR performed histology and immunofluorescence assays. ACL and MJ analyzed the data. ACL and MJ wrote the paper with input from all authors. AML and DFC supervised the project.

Literature Cited

Bastos, H., Lassalle, B., Chicheportiche, A., Riou, L., Testart, J., Allemand, I. and Fouchet, P. (2005). Flow cytometric characterization of viable meiotic and postmeiotic cells by Hoechst 33342 in mouse spermatogenesis. *Cytometry A,* **65**: 1, 40-49.

Bryant, J.M., Meyer-Ficca, M.L., Dang, V.M., Berger, S.L. and Meyer, R.G. (2013). Separation of spermatogenic cell types using STA-PUT velocity sedimentation. *J Vis Exp,* 80,

Carelli, F.N., Hayakawa, T., Go, Y., Imai, H., Warnefors, M. and Kaessmann, H. (2016). The life history of retrocopies illuminates the evolution of new mammalian genes. *Genome Res,*

Castaneda, J., Genzor, P., van der Heijden, G.W., Sarkeshik, A., Yates, J.R., 3rd, Ingolia, N.T. and Bortvin, A. (2014). Reduced pachytene piRNAs and translation underlie spermiogenic arrest in Maelstrom mutant mice. *EMBO J,* **33**: 18, 1999-2019.

Chang, Y.F., Lee-Chang, J.S., Panneerdoss, S., MacLean, J.A., 2nd and Rao, M.K. (2011). Isolation of Sertoli, Leydig, and spermatogenic cells from the mouse testis. *Biotechniques,* **51**: 5, 341-342, 344.

Chen, C., Ouyang, W., Grigura, V., Zhou, Q., Carnes, K., Lim, H., Zhao, G.Q., Arber, S., Kurpios, N., Murphy, T.L., Cheng, A.M., Hassell, J.A., Chandrashekar, V., Hofmann, M.C., Hess, R.A. and Murphy, K.M. (2005). ERM is required for transcriptional control of the spermatogonial stem cell niche. *Nature,* **436**: 7053, 1030-1034.

Chowdhury, R., Bois, P.R., Feingold, E., Sherman, S.L. and Cheung, V.G. (2009). Genetic analysis of variation in human meiotic recombination. *PLoS Genet,* **5**: 9, e1000648.

Ellwart, J.W. and Dormer, P. (1990). Vitality measurement using spectrum shift in Hoechst 33342 stained cells. *Cytometry,* **11**: 2, 239-243.

Gan, H., Cai, T., Lin, X., Wu, Y., Wang, X., Yang, F. and Han, C. (2013). Integrative proteomic and transcriptomic analyses reveal multiple post-transcriptional regulatory mechanisms of mouse spermatogenesis. *Mol Cell Proteomics,* **12**: 5, 1144-1157.

Gaysinskaya, V. and Bortvin, A. (2015). Flow cytometry of murine spermatocytes. *Curr Protoc Cytom,* **72**: 7 44 41-47 44 24.

Gaysinskaya, V., Soh, I.Y., van der Heijden, G.W. and Bortvin, A. (2014). Optimized flow cytometry isolation of murine spermatocytes. *Cytometry A,* **85**: 6, 556-565.

Getun, I.V., Torres, B. and Bois, P.R. (2011). Flow cytometry purification of mouse meiotic cells. *J Vis Exp,* 50,

Getun, I.V., Wu, Z.K., Khalil, A.M. and Bois, P.R. (2010). Nucleosome occupancy landscape and dynamics at mouse recombination hotspots. *EMBO Rep,* **11**: 7, 555-560.

Gonzalez, R. and Dobrinski, I. (2015). Beyond the mouse monopoly: studying the male germ line in domestic animal models. *ILAR J,* **56**: 1, 83-98.

Good, J.M., Giger, T., Dean, M.D. and Nachman, M.W. (2010). Widespread over-expression of the X chromosome in sterile F(1)hybrid mice. *PLoS Genet,* **6**: 9, e1001148.

Laiho, A., Kotaja, N., Gyenesei, A. and Sironen, A. (2013). Transcriptome profiling of the murine testis during the first wave of spermatogenesis. *PLoS One,* **8**: 4, e61558.

Lassalle, B., Bastos, H., Louis, J.P., Riou, L., Testart, J., Dutrillaux, B., Fouchet, P. and Allemand, I. (2004). 'Side Population' cells in adult mouse testis express Bcrp1 gene and are enriched in spermatogonia and germinal stem cells. *Development,* **131**: 2, 479-487.

Liu, Y., Niu, M., Yao, C., Hai, Y., Yuan, Q., Liu, Y., Guo, Y., Li, Z. and He, Z. (2015). Fractionation of human spermatogenic cells using STA-PUT gravity sedimentation and their miRNA profiling. *Sci Rep,* **5**: 8084.

Margolin, G., Khil, P.P., Kim, J., Bellani, M.A. and Camerini-Otero, R.D. (2014). Integrated transcriptome analysis of mouse spermatogenesis. *BMC Genomics,* **15**: 39.

McCarrey, J.R. (2015). EPIGENETICS. The epigenome--a family affair. *Science,* **350**: 6261, 634-635.

Namekawa, S.H., Park, P.J., Zhang, L.F., Shima, J.E., McCarrey, J.R., Griswold, M.D. and Lee, J.T. (2006). Postmeiotic sex chromatin in the male germline of mice. *Curr Biol,* **16**: 7, 660-667.

Rathke, C., Baarends, W.M., Awe, S. and Renkawitz-Pohl, R. (2014). Chromatin dynamics during spermiogenesis. *Biochim Biophys Acta,* **1839**: 3, 155-168.

Rodriguez-Casuriaga, R., Folle, G.A., Santinaque, F., Lopez-Carro, B. and Geisinger, A. (2013). Simple and efficient technique for the preparation of testicular cell suspensions. *J Vis Exp,* 78,

Roig, I., Dowdle, J.A., Toth, A., de Rooij, D.G., Jasin, M. and Keeney, S. (2010). Mouse TRIP13/PCH2 is required for recombination and normal higher-order chromosome structure during meiosis. *PLoS Genet,* **6**: 8,

Sandhu, L.C., Warters, R.L. and Dethlefsen, L.A. (1985). Fluorescence studies of Hoechst 33342 with supercoiled and relaxed plasmid pBR322 DNA. *Cytometry,* **6**: 3, 191-194.

Schlatt, S. and Ehmcke, J. (2014). Regulation of spermatogenesis: an evolutionary biologist's perspective. *Semin Cell Dev Biol,* **29**: 2-16.

Shimizu, Y., Motohashi, N., Iseki, H., Kunita, S., Sugiyama, F. and Yagami, K. (2006). A novel subpopulation lacking Oct4 expression in the testicular side population. *Int J Mol Med,* **17**: 1, 21-28.

Smith, P.J., Nakeff, A. and Watson, J.V. (1985). Flow-cytometric detection of changes in the fluorescence emission spectrum of a vital DNA-specific dye in human tumour cells. *Exp Cell Res,* **159**: 1, 37-46.

Soumillon, M., Necsulea, A., Weier, M., Brawand, D., Zhang, X., Gu, H., Barthes, P., Kokkinaki, M., Nef, S., Gnirke, A., Dym, M., de Massy, B., Mikkelsen, T.S. and Kaessmann, H. (2013). Cellular source and mechanisms of high transcriptome complexity in the mammalian testis. *Cell Rep,* **3**: 6, 2179-2190.

Steen, H.B. and Stokke, T. (1986). Fluorescence spectra of cells stained with a DNA-specific dye, measured by flow cytometry. *Cytometry,* **7**: 1, 104-106.

Watson, J.V., Nakeff, A., Chambers, S.H. and Smith, P.J. (1985). Flow cytometric fluorescence emission spectrum analysis of Hoechst-33342-stained DNA in chicken thymocytes. *Cytometry,* **6**: 4, 310-315.

Yoshida, S., Sukeno, M., Nakagawa, T., Ohbo, K., Nagamatsu, G., Suda, T. and Nabeshima, Y. (2006). The first round of mouse spermatogenesis is a distinctive program that lacks the self-renewing spermatogonia stage. *Development,* **133**: 8, 1495-1505.

Paper V - Challenges and solutions for ribosome profiling with limited cell numbers: the example of murine male germ cells

(in preparation)

Challenges and solutions for ribosome profiling with limited cell numbers: the example of murine male germ cells

Ana Cristina Lima[1,2,3,4], Nicholas Ho Rui Yuan[4], António Amorim[2,3,5], Donald Franklin Conrad[4,6], Alexandra Manuel Lopes[2,3]

1- Graduate Program in Areas of Basic and Applied Biology (GABBA), Abel Salazar Institute of Biomedical Sciences, University of Porto, Porto, Portugal; 2-Instituto de Investigação e Inovação em Saúde (I3S), Universidade do Porto, Porto, Portugal; 3- Instituto de Patologia e Imunologia Molecular da Universidade do Porto, Porto, Portugal; 4- Department of Genetics, Washington University School of Medicine, St. Louis, MO, USA; 5- Faculty of Sciences of the University of Porto, Porto, Portugal; 6- Department of Pathology & Immunology, Washington University School of Medicine, St. Louis, MO 63110, USA

Corresponding author:

Dr. Alexandra M. Lopes

Population Genetics Group, I3S/IPATIMUP

i3S - Instituto de Investigação e Inovação em Saúde, Universidade do Porto

Rua Alfredo Allen, 208, 4200-135 Porto

Phone: +351 220408800; email: alopes@ipatimup.pt

Abstract

In the omics era, several procedures have been established to investigate regulatory mechanisms at the levels of DNA, RNA and protein. The recent development of ribosome profiling, a technique for deep sequencing of mRNA fragments undergoing translation, has brought many insights into the dynamics and control of translation in different tissues and model organisms. This technology is especially appealing to investigate regulation of intricate developmental processes such as spermatogenesis, though extremely challenging given the limited amount of specific cell-types within the tissue. In this work, we adapted and optimized the standard ribosome profiling protocol to allow for parallel generation of translatome and transcriptome data from limited RNA input samples. Using flow cytometry with Hoechst-33342 staining (Ho-FACS), we isolated 5 major male germ cell populations from the mouse testis and characterized them for the proportion of free mRNA, transcripts undergoing translation and mRNA molecules potentially repressed in ribonucleoprotein complexes. Finally, we describe the applicability of our cDNA library preparation with a commercial kit for small RNA-seq to four germ cell populations isolated by Ho-FACS. We believe our work brings valuable improvements to the already widely used ribosome profiling technique, broadening the scope of its application, especially in the resolution of complex regulatory networks of highly heterogeneous tissues.

Key words: Ribosome profiling; low RNA amount; mouse spermatogenesis; FACS

Introduction

Advances in genome technologies and the continuous reduction of associated costs has been driving a recent drastic increase in omics studies to investigate genetic diversity, gene expression, function of non-coding RNA species, epigenetic signatures and protein repertoires in normal tissues or in the context of disease. Although challenged by the limited amounts of input from isolated specific cell-types, these technologies potentially allow to explore the complexity of highly heterogeneous tissues in a massive scale. Despite the individual useful insights, transcriptomic and proteomic data usually display poor correlations mostly due to regulation of post-transcription, translation and protein degradation (Vogel and Marcotte 2012; Larsson *et al.* 2013; Piccirillo *et al.* 2014).

The topic of translational control has recently gained momentum with the development of ribosome profiling (Ingolia *et al.* 2009). In this technique, ribosome-protected mRNA fragments (RPFs) are generated by nuclease digestion and resulting footprints are isolated and identified by RNA-seq. Ribosome footprint profiles comprise therefore a quantitative catalogue of mRNAs undergoing translation in a cell population/tissue at a given time. The usefulness of this technique is unprecedented, allowing for high resolution genome-wide measurements of ribosome occupancy of mRNAs, translation rates, programmed translation of non-canonical isoforms and non-coding RNAs, and translational responses to stress conditions (Ingolia *et al.* 2009; Ingolia *et al.* 2011; Gerashchenko *et al.* 2012; Ingolia *et al.* 2014). Additionally, treating cells with different drugs to halt translation during elongation (Cycloheximide) or initiation (Harringtonine and Lactomidomycin) generates different footprint profiles that register the usage of alternative initiation sites, upstream open reading frames (uORFs) and the kinetics of translation (Ingolia *et al.* 2011; Ingolia *et al.* 2012; Lee *et al.* 2012). Importantly, the collection of transcriptome data in parallel with ribosome profiling allows for a combined analysis of mRNA transport and translation and for the measurement of translational efficiency as well as of differential translation (Ingolia *et al.* 2009; Ingolia *et al.* 2012; Larsson *et al.* 2013). The contributions brought by ribosome profiling of many tissues/cell types from several model organisms have been enormous, as elegantly summarized by (Jackson and Standart 2015). Although highly efficient, the standard ribosome profiling protocol is a long, labor intensive procedure that calls for a skilled practitioner for its implementation. Importantly, its application to specific *ex vivo* cell-types has been hindered by the necessity of high quantities of RNA isolated from ribosomes (>2µg) to generate cDNA libraries.

The unique regulatory complexity driving the development of male germ cells during spermatogenesis is a very appealing system to be resolved by ribosome profiling. Indeed,

ribosome profiling of MAEL knockout mice testis has revealed translational defects likely resulting from disturbances in the key players of mRNA silencing, piRNAs and MIWI (Castaneda *et al.* 2014). Moreover, translational control during stages of repressed transcription, such as meiotic prophase and from mid-spermiogenesis onwards (Kierszenbaum and Tres 1978; Hecht 1998; Bradley *et al.* 2004; Wang *et al.* 2005; Rolland *et al.* 2008), calls for a stage-wise translatome/transcriptome comparison in different germ cell-types.

Here, we investigate the applicability of ribosome profiling to different germ cell-types isolated from murine testis by fluorescence-activated cell sorting with Hoechst-33342 staining [Ho-FACS; (Bastos *et al.* 2005; Getun *et al.* 2011; Gaysinskaya *et al.* 2014)]. For that, we developed a new protocol using a commercial small RNA-seq kit for library preparation from reduced quantities (100 ng) of ribosome footprints. With this workflow (Fig.1), we can generate cDNA libraries from 300 ng of RPFs for 4 major testicular subpopulations: i) Spermatogonia, ii) primary spermatocytes and iii) spermatids (round and elongating). Although the quantitative power of our approach needs further investigation, it allows for the assessment of differential translation and detection of testis-specific isoforms, providing an overall overview of translational regulation throughout spermatogenesis.

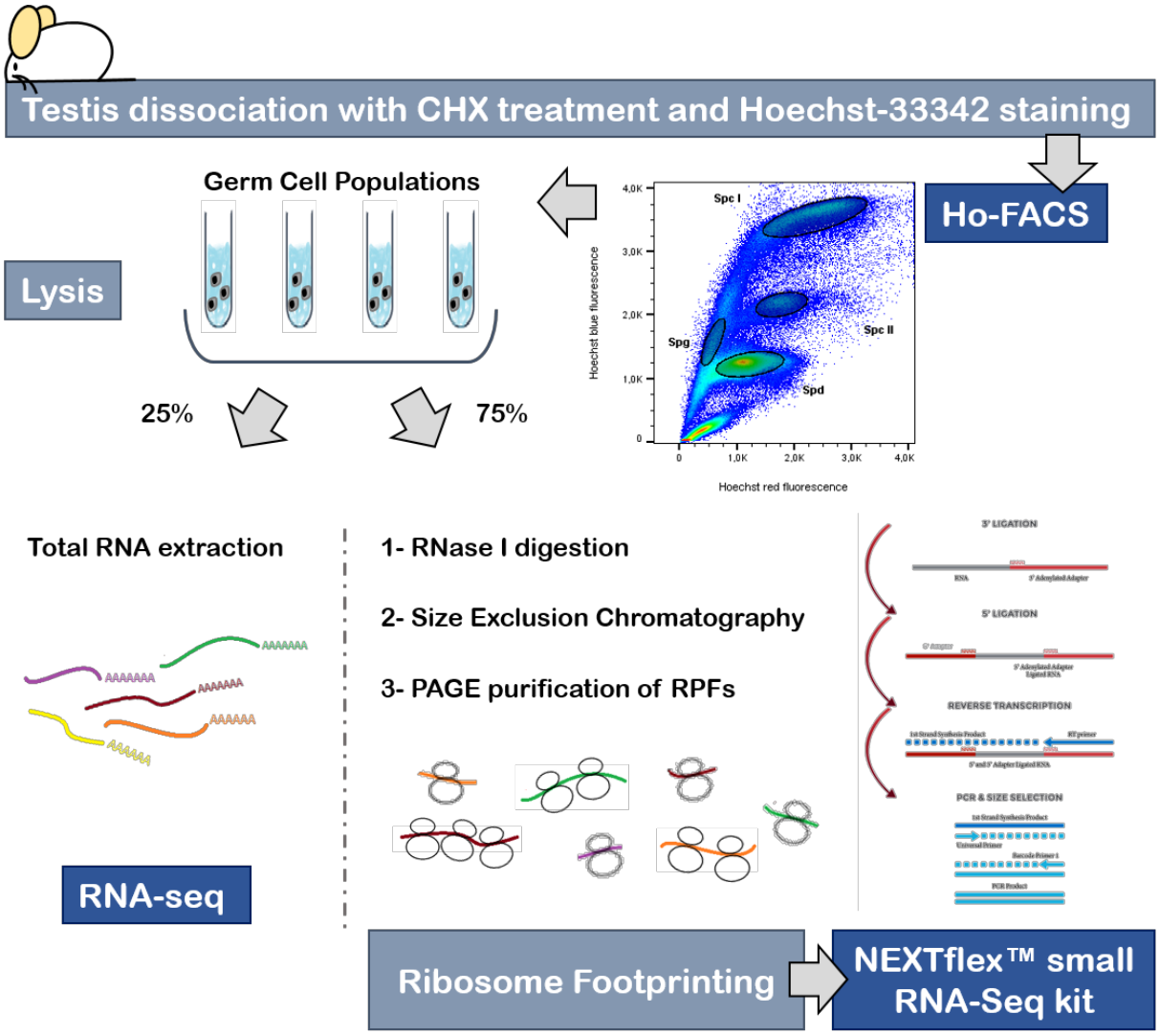


Figure 1. Workflow schematics

Workflow proposed to generate transcriptome and translatome data in parallel from different mouse male germ cells. Mouse testes are collected fresh and enzymatically dissociated. During dissociation, cycloheximide (CHX) is added to stop translation while preserving the cellular state of the ribosomes during elongation and cells are stained with Hoechst-33342 to allow discrimination of different germ cell-types by FACS. Cell lysates from Ho-FACS sorted populations are divided for transcriptome (25%) and ribosome profiling (75%) procedures. For the latter, ribosome footprints are generated by RNase I digestion of lysates, ribosomes are recovered by size exclusion chromatography, RNA is extracted and ribosome protected fragments (RPFs) are purified and size selected by denaturing polyacrilamide gel electroforesis (PAGE). Finally, cDNA libraries are generated from RPFs with NEXTflex™ small RNA-Seq kit using 25% of adapter stock concentration and 18 PCR cycles.

Materials & Methods

Dissociation and Hoechst-33342 staining of murine testes

Wild-type C57BL/6 male mice (Jackson Laboratory) were raised in animal facilities at Washington University in St. Louis and sacrificed after reaching sexual maturity (8 - 12 weeks) in compliance with regulations of the Animal Studies Committee at Washington University in St. Louis. In order to preserve the *in vivo* molecular signature of male germ cells, all solutions for dissociation were prepared fresh, prior to animal sacrifice, and cycloheximide (CHX) was added to the dissociation medium to stop translation elongation. Testes dissociation was performed as described in (Getun *et al.* 2011) with few modifications. Briefly, fresh testes were de-capsulated, rinsed in 1X Phosphate buffered saline (PBS; Thermo Scientific #AM9625) and placed in 15mL tubes containing 1X Dulbecco's Modified Eagle Medium (DMEM; Life Technologies #31053) and CHX (0.1mg/ml; Amresco #94217). Single-cell suspensions were obtained from 2-4 mouse testes by a series of enzymatic digestions performed at 33°C in an orbital shaker by addition of Collagenase I (120 U/mL; Worthington Biochemical #LS004196), DNAse I (1 mg/mL in 50% glycerol; Roche #10104159001) and trypsin (50mg/ml in 10mM HCl; Worthington #LS003708). Fetal Bovine Serum (FBS; Thermo Scientific #10082139) was used to inactive trypsin. At the end, cells were passed through two 40µm 1X DMEM pre-wetted disposable filters (Thermo Scientific, #22363547) to exclude any larger somatic cells still present and cell aggregates resulting from undigested tissue and/or cell death occurring during dissociation. When performing testis dissociation for FACS, cells were stained with Hoechst-33342 (10 mg/mL; Life Technologies, #H3570) and stored on ice protected from light until further processing (less than 1h).

Fluorescence Activated Cell Sorting (FACS) of testicular cell suspensions

Cells were sorted and analyzed by a Beckman Coulter MoFlo Legacy cell sorter using Summit Cell Sorting software, similarly to the descriptions of (Getun *et al.* 2011). Hoechst was excited using a ultra-violet laser and triggered for scatter by a 488nm blue laser. Hoechst Blue and red fluorescence were detected pairing the U.V. laser with a 463/25nm and a 680nm LP band pass filters, respectively. A 555DLP dichroic was also used to distinguish blue from red emission wavelengths. Samples were analyzed using 70-micron nozzle and the sorting flow rate was set to 3000-4000 events/second. A minimum of 500,000 events, were detected before proceeding to gating. Since we did not use Propidium Iodide, we engaged in a sequential cell gating strategy: debris were excluded based on forward (FS) and side scatter (SS) parameters, then

singlets gated by adjusting threshold for forward scatter pulse width and finally red/blue Hoechst fluorescence signal was used to detect populations of spermatogonia, spermatocytes (primary and secondary) and spermatids. To discriminate round from elongating spermatids, we gated for combined high FS and SS or low FS and SS respectively and then for blue/red Hoechst fluorescence as described elsewhere (AC Lima *et al.*, submitted). Each testis was processed for 2.5 hours to recover the maximum number of cells from the full volume of cell suspension. Cells were collected into 1 mL of 1X PBS + 10% FBS in 5 mL polypropylene round-bottom tubes pre-coated with FBS.

Isolation of ribosome-bound and free mRNA molecules

To concentrate the samples and remove dead cells and debris, single-cell suspensions of both FACS sorted and whole testis samples were pelleted by centrifugation (600xg at 4°C for 10 min) and washed with ice-cold 1X PBS supplemented with CHX (0.1 mg/mL). For sorted cell populations, 50 µL of volume was transferred to a 1.5 mL tube for cell fixation in 4% PFA and stored at 4°C in the dark. Cells were then lysed [20 mM Tris·Cl (pH 7.4), 150 mM NaCl, 5 mM MgCl2, 1 mM DTT, 0.1 mg/mL cycloheximide, 1% (vol/vol) Triton X-100 and 25 U/mL Turbo DNase I (2 U/µL; AM2238)] and intracellular material was collected by centrifugation (20.000xg for 10 min at 4°C) and stored at -80°C. 75% of the lysate was used for ribosome footprinting and the remaining 25% for total RNA extraction. All reagents and solutions used here were either purchased RNase-free or prepared with RNase-free water and materials.

Generation of ribosome-protected RNA fragments (RPFs)

Purification of RNA undergoing translation was performed based on the protocols described in Ingolia *et al.* (2012) and ARTseq™ Ribosome Profiling Kit (Epicentre #RPHMR12126). Lysate used for ribosome profiling was treated with RNAse I (100U/µL; Invitrogen #AM2294) for 45 min at RT to generate ribosome footprints and nuclease digestion was stopped by addition of SUPERase-In RNase inhibitor (20 U/mL; Invitrogen #AM2694). Ribosomes were then isolated by size exclusion chromatography with Sephacryl S-400 columns (GE Healthcare #27-5140-01) as described in ARTseq™ Ribosome Profiling Kit protocol guidelines. RNA was extracted with mirVana miRNA extraction kit (Life Technologies #AM1560) following the manufacturer's protocol for organic extraction with acid phenol-chloroform and enrichment for small RNAs (<200nt). For lysates obtained from sorted cells, samples of the same germ cell population obtained in two FACS were pooled during acid phenol-chloroform extraction. During ribosome footprint RNA extraction, the fraction comprising RNA fragments longer than 200 nucleotides

was also collected in parallel, herein referred to as RPFs >200 nt. RNA was quantified using Nanodrop and precipitated [Glycoblue (15 mg/mL; Invitrogen #AM9515); 3M sodium acetate (pH 5.5; Invitrogen #AM9740); 100% isopropanol (Fisher Scientific #BP2618500)] overnight at -80°C. Precipitated RNA was collected by centrifugation (20,000xg for 30 min at 4°C), washed with 80% EtOH and ressuspended in 10 mM Tris (pH 8.0; Invitrogen; AM9855G).

Size selection and purification of RPFs

To discriminate true monosome-protected RNA fragments from those within polysomes or other protein complexes, we performed denaturing (15% TBE-Urea; BioRad 456-6055) polyacrylamide gel electrophoresis (PAGE) for the RPFs <200 nt samples, as described in Ingolia *et al.* (2012). The expected size range of monosome-bound RNA (26-34 nt; herein monosome fraction) was detected using a combination of upper [5`-AUGUACACGGAGUCGAGCUCAACCCGCAACGCGA-(Phos)-3`] and lower [5`-AUGUACACGGAGUCGACCCAACGCGA-(Phos)-3`] oligo markers (IDT). Bands comprising the monosome fraction were excised using the oligo markers as reference, bands with larger sizes (labeled as polysome fraction) were also cut (Fig.6), and RNA was recovered following the overnight RNA gel extraction protocol. Precipitated RNA was collected by centrifugation (20,000xg for 30 min at 4°C), washed with 80% EtOH and ressuspended in RNase-free water.

Total RNA extraction

RNA was extracted from total cell lysates with mirVana miRNA extraction kit (#AM1560; Life Technologies) according to the manufacturer's protocol for organic extraction with acid phenol-chloroform and total RNA extraction. For lysates obtained from sorted cells, two samples of the same FACS sorted population were pooled during acid phenol-chloroform extraction. RNA was quantified using Nanodrop and precipitated [Glycoblue (15 mg/mL; Invitrogen #AM9515); 3M sodium acetate (pH 5.5; Invitrogen #AM9740); 100% isopropanol (Fisher Scientific (#BP2618500,)] overnight at -80°C. Precipitated RNA was collected by centrifugation (20,000xg for 30 min at 4°C), washed with 80% EtOH, ressuspended in RNase free water and stored at -80°C.

Preparation and optimization of cDNA libraries from RPFs

Libraries were prepared using NEXTflex™ small RNA-Seq kit (#5132-03, BIOO SCIENTIFIC®), which is optimized to generate libraries from 1-10 µg total RNA or purified small RNA from 1-10 µg total RNA. Experiments were performed according to the manufacturer's

guidelines for small RNA starting material and 18 cycle PCR reactions with the following exceptions:

-Adapter tests: 100ng of RPFs recovered from whole testis were used to evaluate the performance of library preparation with variable adaptor concentrations. Stock solutions for 5'adaptor, 3'adaptor, reverse transcription primer and PCR primers provided by the library prep kit were diluted in RNase-free water to final concentrations of 100%, 75%, 50% and 25%. To remove any gDNA contamination, all samples were treated with Turbo DNase I (2 U; 1X TurboDNase Buffer; #AM2238, Life Technologies) for 30min at 37°C prior to library preparation. To increase efficiency, 3' ligation was performed overnight at 22°C.

- Nuclease digestion and PCR amplification tests: Cell lysate of one testis was divided in 2 and treated with RNase I for 1h at room temperature or 37°C for footprint generation. 100 ng of RPFs recovered from both digestion conditions were either PAGE purified for size selection or used directly for library preparation without size selection. Samples were treated with Turbo DNase I (2 U; 1X Turbo DNase Buffer; #AM2238, Life Technologies) for 30min at 37°C prior to library preparation, which was performed with 25% concentration of adaptors and primers. cDNA generated from PAGE purified RPFs was amplified with 22 and 25 cycles, whereas cDNA of RPFs without size selection was amplified with 18 PCR cycles.

- Recovery of cDNA libraries: After size selection with TBE-PAGE, cDNA libraries were extracted from the bands following the overnight DNA gel extraction described in Ingolia *et al.* (2012).

For FACS sorted murine germ cell subpopulations, as input material we used RNA extracted from both monosome and polysome fractions separated by denaturing PAGE. cDNA libraries were generated with [25%] adaptors and primers after DNase treatment.

Prior to sequencing, all libraries were quantified with 2100 BioAnalyzer (#G2940CA; Agilent Technologies) using Agilent High Sensitivity DNA Kit (#5067-4626; Agilent Technologies).

Generation of ribosome profiling control data

Translatome data was generated from one mouse testis to serve as positive control, following the protocol described in Ingolia *et al.* (2012). Gel extractions and RNA precipitations were always carried overnight to maximize recovery. 8 and 10 cycles were used for PCR amplification. Variations from the protocol were as follows: Ribosomes were isolated by size exclusion chromatography with Sephacryl S-400 columns (GE Healthcare; 27-5140-01) as described in ARTseq™ Ribosome Profiling Kit protocol guidelines; RNA extractions were performed using mirVana miRNA extraction kit (AM1560; Life Technologies) according to the

manufacturer's protocol for organic extraction with acid phenol-chloroform and total RNA extraction.

Sequencing and data analysis

To evaluate the efficacy of the cDNA library preparation from RPFs using the NEXTflex™ small RNA-Seq kit, we sequenced the libraries generated using different concentrations of RNA input (RP_Kit 10 µg; RP_Kit 1 µg; RP_Kit 100 ng and RP_Kit 10 ng). Additionally, we sequenced the library obtained following the original ribosome profiling protocol (RP_Ingolia) as a control. Paired-end sequencing of cDNA libraries was performed in Hi-Seq or Mi-Seq Illumina platforms with 2X 101 or 2X150 protocols and different read depths. Raw sequencing reads were mapped to the mm10 assembly of the *Mus musculus* genome using bwa mem with default parameter settings (Li and Durbin 2009). The distribution of reads mapping to 5'-UTRs, 3'-UTRs, CDS and introns was calculated using the script read_distribution.py from the RNASeQC package (DeLuca *et al.* 2012). Additionally, we estimated the insert size distribution to confirm that we the RNA sequenced had the expected size of monosome-bound RPFs, which is 26-34 nt. We also evaluated the degree of rRNA contamination in our samples by estimating the percentage of raw reads mapping to rRNA sequences. A table with summary statistics from the sequencing data can be found in the supplementary material (Supplementary Table 2). Finally, we compared the datasets obtained here for different RNA inputs or with the dataset generated by (Castaneda *et al.* 2014) for wt mice.

Results

Successful isolation of total RNA and RPFs from Ho-FACS sorted male germ cells

Until now ribosome profiling experiments have been only performed in whole testis. Given that RNA molecules bound to ribosomes (ribosome protected fragments; RPFs) represent only a very small fraction of the total RNA present in a cell, our first step was to effectively quantify the amount of total RNA we could recover from the different cell-types of mouse testis. We used different gating strategies during FACS to isolate spermatogonia (Spg), spermatocytes (primary: Spc I and secondary: Spc II) and spermatids (Spd), further resolved into round (rSpd) and elongating (eSpd), from the mouse testis (Fig.2A-C). From a total of 9 experiments, we

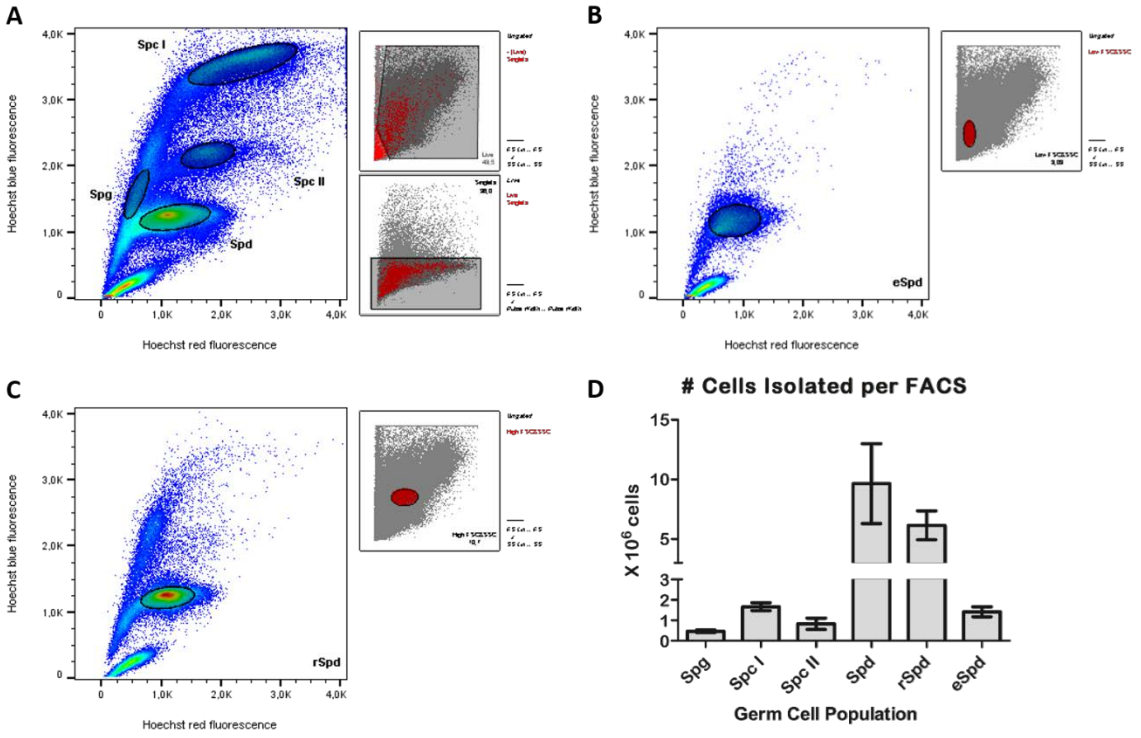


Figure 2. Ho-FACS isolation of male germ cell populations

Mouse testicular cell suspensions stained with Hoechst-33342 were used to isolate different germ cell populations by FACS with different gating strategies. Gating for live single-cells and plotting the function of blue and red Hoechst fluorescence allowed us to discriminate Spg, Spc I, Spc II and Spd (A). To resolve the Spd population into rSpd and eSpd we gated based on combined forward (FS) and side scatter (SS) paramenters, representatives of cell size, shape and complexity. eSpd are smaller, more elliptical cells with more compact chromatin showing lower FS & SS values (B) while rSpd present higher FS & SS values (C). Estimation of the average number of cells sorted per population reflects the proportion of each germ cell type present in the mouse seminiferous tubules (D). The combined cell numbers of individually gated rSpd and eSpd (7.57 x10^6) add to slightly lower values than total numbers of Spg (9.65 x10^6), reflecting the stringency of the gates applied. Spg: spermatogonia; Spc I: primary spermatocytes; Spc II: secundary spermatocytes; Spd: Spermatids; rSpd: round spermatids; eSpd: elongating spermatids. Smaller plots illustrate the parental gates used to generate the main red/blue Hoechst fluorescence plots (A-C). Since we can only sort 4 populations per FACS session the average number of cells was estimated based on 9 FACS for Spg and Spc I, 7 for rSpd and eSpd and 2 for Spc II and Spd (D). FACS plots were obtained using FlowJo® software v10 (Tree Star Inc.). Histogram was generated with GraphPad Prism (version 5.02 for Windows, GraphPad Software, San Diego California USA, www.graphpad.com), plotting the calculated mean values with standard deviation.

estimated that we can FACS sort on average a minimum of ~0.5 x10^6 spermatogonia, the less abundant germ cell population, and a maximum of ~6.1 x10^6 round spermatids, the most

numerous cell type in mouse testis (Fig.2D and Supplementary Table 1). We then used cell lysates from each of these sorted populations to recover both total RNA and RPFs (Methods). As expected, nuclease digestion generated RPFs of variable sizes with a generalized pattern consistent to the one described by Ingolia *et al.* (2012) (Fig.2A-B). Figure 3C shows the average concentration of RPFs and total RNA estimated for each cell population, normalized to the cell number and percentage of cell lysate used for each RNA fraction (75% and 25%, respectively). Interestingly, these measurements provide a naïve comparison of the dynamics of gene expression and translation throughout spermatogenesis. The most striking observation is the reduction of total RNA and RPFs in spermatids compared to spermatogonia and spermatocytes. Also, when gated together, the relative percentage of different RNA fractions in Spd resembles the one registered for rSpd, likely due to the higher proportion of rSpd within the seminiferous tubules (Mays-Hoopes *et al.* 1995) Further separation of two spermatid subpopulations provided a higher resolution of the RNA dynamics in this stage , with both rSpd and eSpd presenting a higher transcription than translation rate (RPFs <200), and clearly exemplifies the need of analyzing different cell-types individually. Highest values are detected for transcription in spermatogonia and active translation in secondary spermatocytes. Also, the proportion of RPFs >200nt suggests that translational regulation of mRNAs occurs mainly in Spg and eSpd.

Considering the minimum average number of cells (Fig.2D) and amount of RNA recovered (Fig.3C), we estimated the expected yield of different RNA species we can obtain on average for each population from one FACS session (Supplementary table 1). Importantly, these results indicate that we are able to consistently recover at least ~300ng of both total RNA and RPFs from the same sample for all FACS sorted germ cell populations. This information was used to set the lower limit threshold of starting material for library preparation.

Library preparation of RPFs from low RNA input samples

Next, we evaluated the efficiency of the NEXTflex™ small RNA-Seq kit (BIOO SCIENTIFIC®) in the preparation of cDNA libraries from RPFs. Briefly, library preparation with this kit comprises two steps of adapter ligation (3' and 5'), a first strand cDNA synthesis with a reverse transcription primer annealing to the 3´ adenylated adapter followed by PCR amplification with barcoded reverse primers. During the protocol, two clean-up steps with magnetic beads (Agencourt AMPure XP, Beckman Coulter) reduce the accumulation of unligated adapters to help prevent the formation of adapter dimers during PCR amplification. At the end, PCR products are purified by non-denaturing PAGE to separate adapter dimers (~130bp) from cDNA

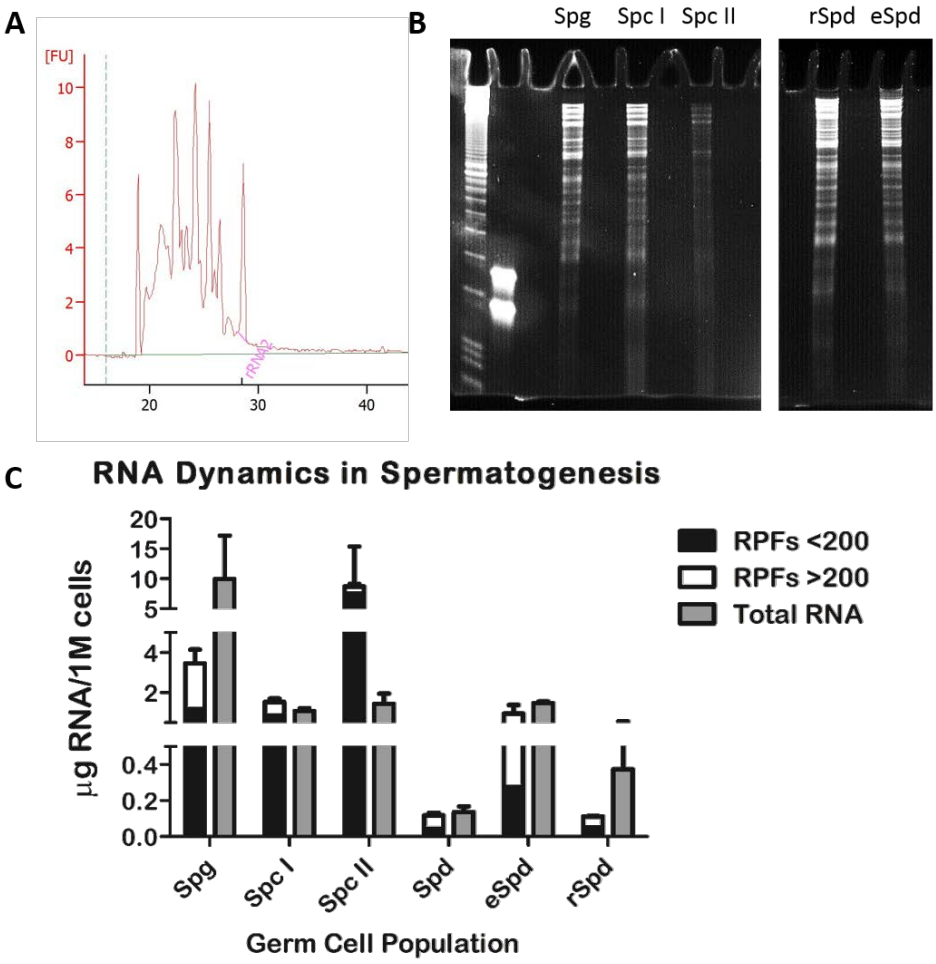


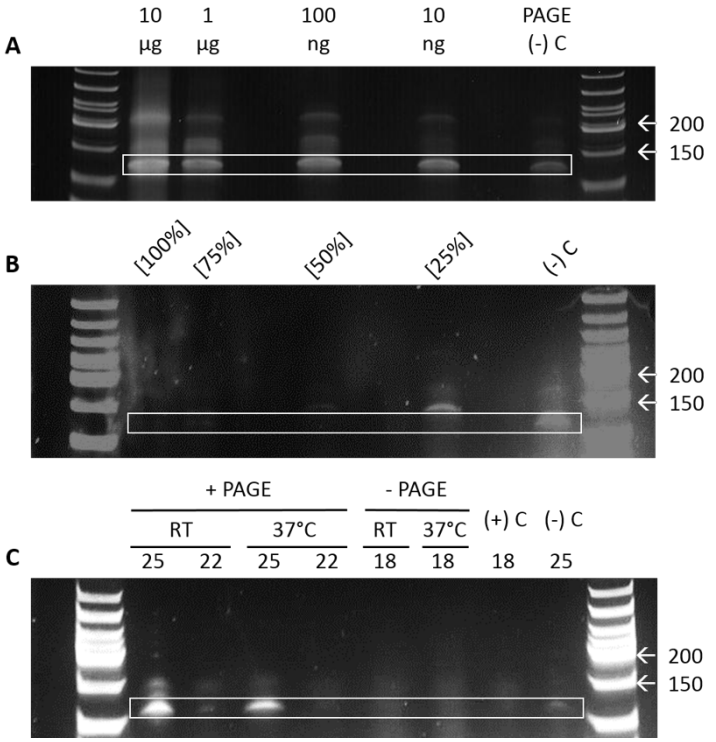
Figure 3. Isolation of RPFs and total RNA from different male germ cell types isolated by Ho-FACS.

Total RNA and RPFs generated by RNase I digestion were extracted from Ho-FACS sorted male germ cell populations using 25% and 75% of cell lysates, respectively. During RNA extraction, RPFs larger and smaller than 200 nts were recovered seperately (RPFs >200 and RPFs <200, respectively) while total RNA only comprises RNA molecules larger than 200 nt. As expected, ribosome footprints (RPFs <200) comprise a range of RNA fragments of different sizes, as confirmed by the profiles obtained by Agilent BioAnalyzer (A) and by denaturing polyacrilamide gel electrophoresis (PAGE; B). Total RNA and RPFs were quantified and normalized to the volume of lysate and cell number to estimate the amount of RNA present in 10^6 cells (C). This estimation provides a generalized idea of transcription and translation rates in different male germ cells. Spg: spermatogonia; Spc I: primary spermatocytes; Spc II: secundary spermatocytes; Spd: spermatids; rSpd: round spermatids; eSpd: elongating spermatids.BioAnalyzer traces were obtained running 1µL of sample in Agilent RNA 6000 Pico chips. PAGE was performed with 15% TBE-Urea gels (Bio-rad) and stained with SYBR gold (Invitrogen) for visualization. A 10-bp DNA ladder (Lane 1 in B) and 26-34 nt oligomarkers (Lane 2 in B) were used as size references. Quantifications were obtained from at least two experimental replicates for each cell type and histogram was generated with GraphPad Prism (version 5.02 for Windows, GraphPad Software, San Diego California USA, www.graphpad.com), plotting the mean values with standard deviation.

libraries (~150bp). Although the kit was designed to the preparation of libraries generated from small RNAs, cDNAs obtained from RPFs are expected to produce a similar size of PCR amplicons since their length (26-34 nts) falls within the size range of small RNAs (18-45 nts).

We first tested the protocol's performance with a range of RPFs concentrations obtained from whole mouse testis (See methods; Fig. 4A). As expected, efficiency of library preparation decreases with the amount of RNA input as illustrated by the intensity of cDNA library bands (~150bp). The negative control sample was obtained from gel extraction of an empty gel lane during size selection of RPFs and the presence of very faint bands >130bp suggests contamination across lanes. Surprisingly, bands which should represent adapter dimers (~130bp; white box in Figure 4A) show a similar correlation to the quantity of RNA input, with the weakest signal detected in the negative control. Given that RPFs comprise a range of sizes, this suggests that the length of some cDNAs might overlap with the expected band of adapter dimers. Deep-sequencing of the bands with ~150bp (Supplementary table 2) shows greater percentages of unique reads mapping to the mouse genome reference sequence (mm10) for samples with higher RPFs input. Sequence analysis of the unmapped reads indicates they correspond to adapter dimers, confirming that the concentration of adapters used for library preparation greatly influences the quality of sequencing data and should be adjusted to the amount of starting material.

Given these observations, we tested the use of different adapter concentration in the preparation of libraries from 100ng of RPFs (Fig.4B). It seems that the ideal RNA/adapter ratio is found using 25% of adaptor stock concentration, as indicated by the strongest cDNA band (~150bp). The absence of stronger cDNA bands with higher adapter concentrations was unexpected and could have resulted from a combined effect of RNA degradation during DNase treatment at 37°C and lower ligation efficiency due to unbalanced adapter/RNA ratios. Next, we attempted to optimize the efficiency of this workflow by adjusting conditions so that we could increase recovery of RPFs from the same initial amount of sample and/or boost amplification of cDNA libraries. For that we: i) evaluated the effect of temperature incubations during nuclease digestion on the yield of monosome footprints recovered from cell lysates of whole testis, ii) skipped the step of RPFs purification by denaturing PAGE and iii) increased the number of PCR amplification cycles from the 18 recommended in the kit (Methods). Nuclease digestion at 37°C does not appear to generate more monosome footprints and a reduction in concentration of recovered RPFs suggests that it may degrade RNA, resulting in an overall decrease in yield. (Supp. Fig.1). This is confirmed by the fainter bands detected after library PCR with either 22 or 25 cycles (Fig.4C). As expected, 25 PCR cycles seem perform better



	RPFs	RNase I digestion	PAGE Size selection	DNase treatment	[Adapter]	3´ligation	PCR cycles
A	Variable	RT	monosomes	NO	100%	2h	18
B	100 ng	RT	monosomes	YES	Variable	Overnight	18
C	100 ng	Variable	Variable	YES	25%	2h	Variable

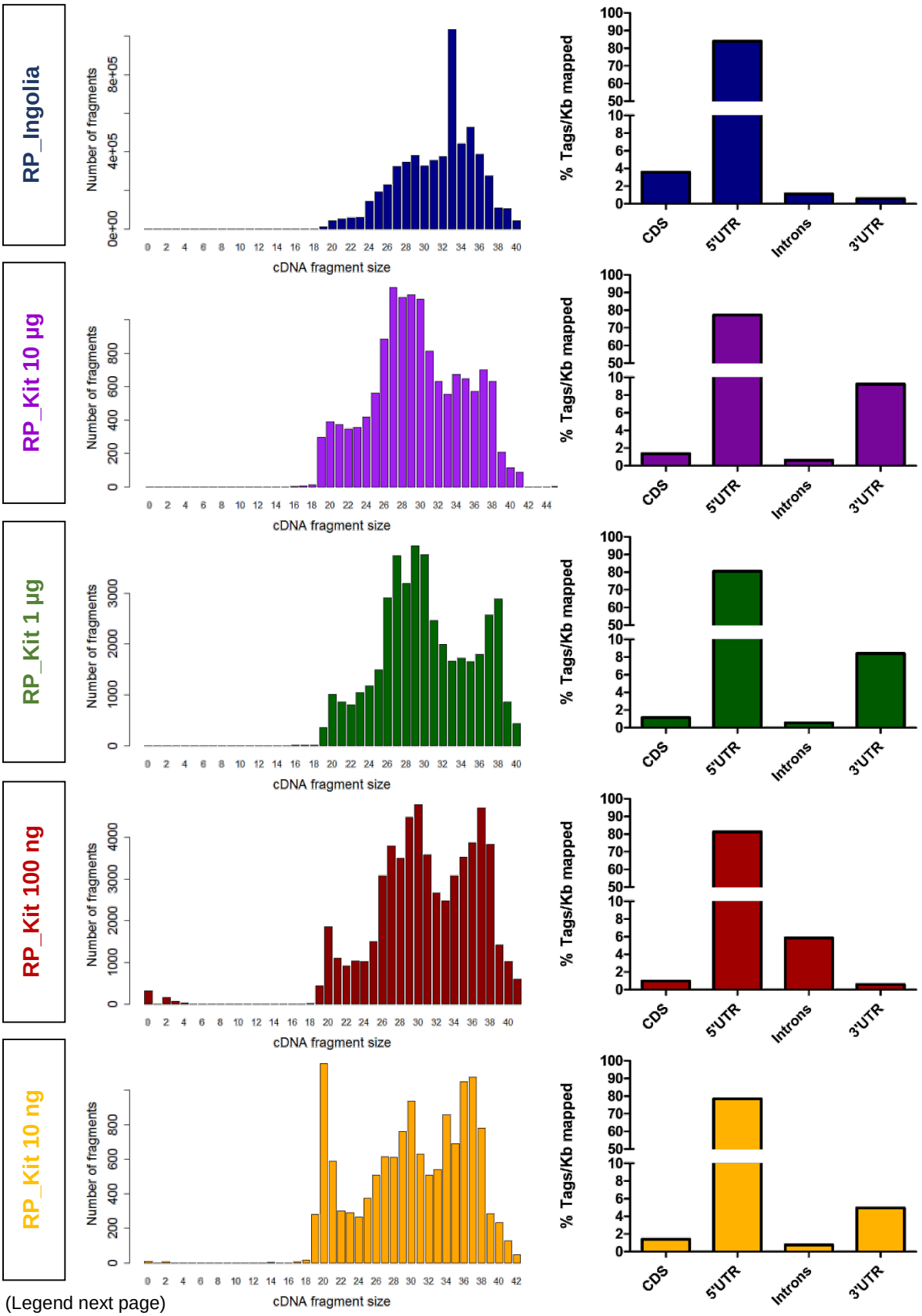
Figure 4. Optimization of cDNA library preparation.

The applicability of NEXTflex™ small RNA-Seq kit to the generation of cDNA libraries from RPFs obtained from whole mouse testis was evaluated using different concentrations of starting material (A) or with 100 ng of RPFs: (B) decreasing concentrations of adapters; and (C) performing digestion at room temperature (RT) or 37°C with a variable number of amplification cycles in size selected RPFs (~26-34 nt; + PAGE) and RPFs <200 nt not PAGE purified (- PAGE). To prevent contamination with gDNA, a DNase treatment was performed prior to library preparation (B-C). General conditions are summarized in (D). Adapter dimers are predicted to be ~130 bp and are maked by white boxes. As expected, higher RNA concentrations yield more cDNA, though even 10ng of RPFs show PCR product (A). For 100ng of RPFs, 25% of adapter concentration seems to result in higher efficiency in ligation as demonstrated by the intensity of the cDNA bands (B). Also, nuclease digestion at room temperature (RT) and 25 PCR cycles seem to generate more cDNA libraries (C). Higher intensity of bands ~130 bp in comparison to the negative controls suggests that some cDNA fragments might overlap in size with adapter dimers (A and C).PAGE was performed with 5% or 10% TBE gels (Bio-rad) and stained with SYBR gold (Invitrogen) for visualization. A ready to load Low MW Ladder provided with the kit was used for size reference.

with cDNA libraries showing bands ranging from 130-150bp. Again, negative control shows only a faint band of putative adapter dimers (~130bp) indicating that some cDNA fragments might overlap in size with adapter dimers. Given the cDNA band observed using 25% of adapter concentration and 18 PCR cycles (Fig.4B), the lack of strong defined cDNA bands for both RPFs without size selection (Fig.4C – PAGE samples) and positive control suggests higher efficiency of overnight 3'-ligation.

In order to obtain a generalized idea of the effect of RNA input in the library preparation, we evaluated the quality of our ribo-seq data in terms of summary statistics of % of reads mapping to the reference sequence of the mouse genome and rRNA contamination (Supplementary table 2), as well as the size distribution of the fragments sequenced and in which regions of the gene bodies they were mapping (Fig.5). This analysis was also performed for a positive control dataset obtained following the original ribosome profiling protocol [RP_Ingolia; (Ingolia et al. 2012)]. This was important since the only published dataset of wt mouse testicular translatome (Castaneda *et al.* 2014) was generated from ribosomes recovered by ultracentrifugation, whereas we used size exclusion chromatography (SEC) instead. For the samples prepared using the commercial kit 8-47.2% of reads mapped to mm10, with the lowest value observed for the library prepared from 10ng RPFs, whereas the for the RP_Ingolia sample 66% of reads were mapped (Supplementary table 1). Interestingly, the Kit sample with higher percentage of mapped reads (47.2%) was prepared from 1µg of RPFs suggesting that this RNA concentration might represent the best RNA/adapter ratio when using the original kit settings. Importantly, the fragments sequenced show the expected size range for all libraries prepared, and the majority of the reads where mapping to 5-UTRs, as expected due to CHX treatment (Fig.5).

Comparison of the RP_Ingolia sample with that of (Castaneda et al. 2014) generated from whole testis of wild-type (P21 Mael$^{129+/+}$) mice shows a good overall correlation (r^2=0.56; Fig.6), indicating that using size exclusion chromatography as an alternative to ribosome isolation by ultracentrifugation does not seem to significantly alter the profile of the data. Moreover, this represents the minimum achievable concordance since differences in mapping methods, gene models and applied filters could influence the overall correlation between the datasets. When comparing transcript data generated by different library preparation protocols, we observed a poor correlation between the datasets, possibly influence, at least to some extent, by the excess of reads mapping to 3'-UTRs detected in the samples prepared with the commercial kit (Fig.5). Importantly, the ribo-seq data generated with 10 µg and 100 ng of RPFs show a high correlation (r^2=0.81), suggesting reduced concentrations of RNA input do not affect the type of transcripts captured using the kit.



(Legend next page)

Figure 5. Distribution of fragment sizes and mapping across gene bodies.

The size of the cDNA fragments sequenced was estimated for each sample and shows de expected length of monosome-bound mRNA fragments (left panel). The majority of these cDNAs mapped to 5'-UTRs, which was expected since CHX treatment to halt translation results in an overrepresentation of reads at these regions (right panel). When comparing different protocols for library preparation, it appears that the samples prepared with the commercial kit have an increase in reads mapping to 3'-UTRs and a small reduction in the CDS regions. RP_Kit 10 presented outliers with very low counts of insert sizes ranging from 40-150 nt that were excluded from the plot. Plots are color-coded in relation to the dataset they represent and were generated using R software (R Development Core Team 2012) and GraphPad Prism (version 5.02 for Windows, GraphPad Software, San Diego California USA, www.graphpad.com).

Library preparation of RPFs from Ho-FACS sorted male germ cells

We then tested this protocol in FACS sorted male germ cell populations, using 25% of adapter concentration, DNase treatment and 25 cycles of library PCR amplification. Since we can only sort 4 samples at a time, we selected the populations predicted to present greater translational regulation: Spg, Spc I, rSpd and eSpd. Different concentrations of RPFs were obtained pooling samples from two FACS (Fig.7A) and selected for size. Both monosome (red box in Figure 6B) and polysome (green box in Figure 6B) fractions were recovered for each cell type. PAGE of the cDNA libraries generated for these samples (Fig.7C) shows very faint bands for RPFs recovered from monosomes when compared to the stronger signal of bands for the polysome fraction. Interestingly, these bands are ~130bp in size and appear to be depleted of adapter dimers, as indicated by the variation of intensity between samples as well as the absence of a band with similar size in the negative control lane. These results confirm our previous observations that cDNA libraries from RPFs using the NEXTflex™ small RNA-Seq kit might overlap in size with the expected adapter dimers. Also, the intense smear observed in the positive control is indicative of RNA degradation and confirms the observations in Figure 4.

Overall, our results suggest that, upon optimization, it is possible to generate cDNA libraries from RPFs isolated from different germ cell populations using the NEXTflex™ small RNA-Seq kit. The optimal experimental conditions may rely on the use of 25% of adapters and a maximum of 18 PCR cycles to avoid biased over-amplification of specific RPFs, but need further investigation.

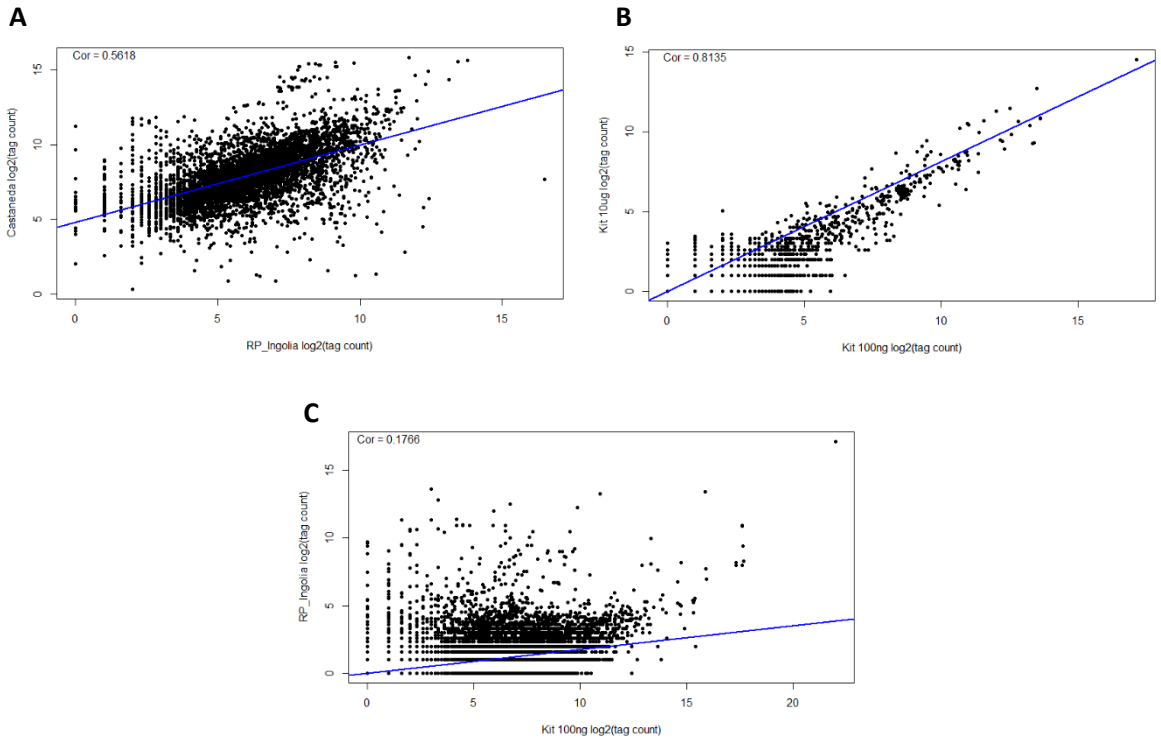


Figure 6. Comparison of ribo-seq datasets generated with different protocols.

In general, it appears that data generated using the same protocol for library preparation (A-B) correlates well, whereas there seems to be a fair difference between the transcripts captured by different methods (C). Importantly, ribo-seq data generated from high or low input of RPFs shows a very good correlation suggesting that the amount of starting material is not affecting the type of transcripts detected. The tag count/transcript was estimated for each sample as used to compare the datasets generated here or obtained from Castaneda *et al.* (2014), in terms of transcripts detected. Plots were generated using R software (R Development Core Team 2012).

Discussion

Flow cytometry with Hoechst-33342 staining (Ho-FACS) allows the isolation of different populations of testicular germ cells (Bastos *et al.* 2005; Getun *et al.* 2011; Gaysinskaya *et al.* 2014) that are suitable for high-throughput studies (Chalmel *et al.* 2007; Fallahi *et al.* 2010). Indeed, murine germ cell populations isolated by Ho-FACS yield a minimum average of ~300ng of ribosome-protected mRNA fragments (RPFs) and total RNA from as little as $0.5x10^6$ cells (Fig.2-3 and Supp. Table.1). During ribosome footprinting, close proximity of ribosomes within an active polysome can block RNase access to the RNA molecule and prevent its

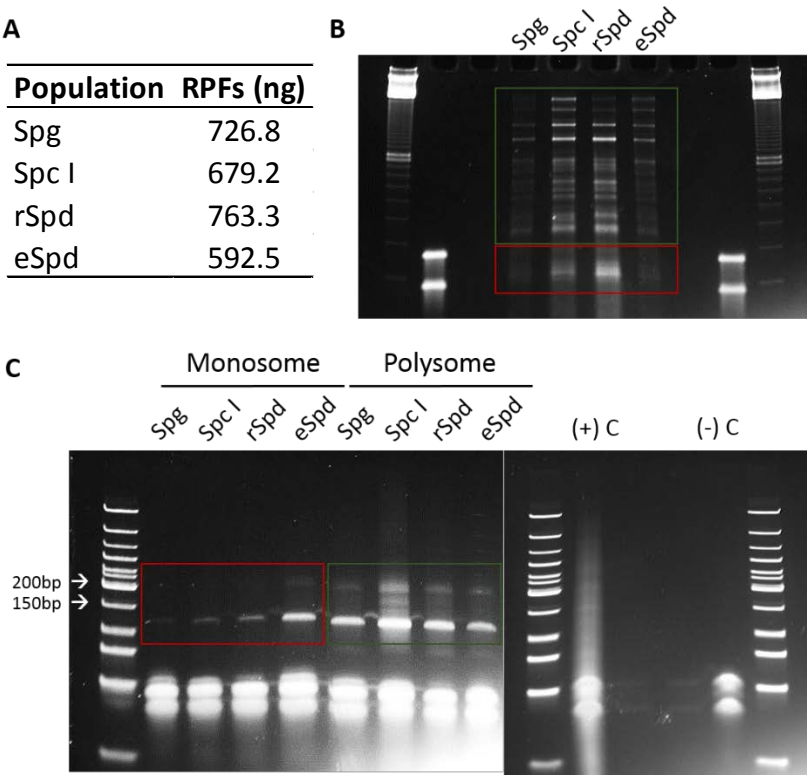


Figure 7. Library preparation of RPFs from different murine male germ cells.

Ribosome footprints were obtained from Ho-FACS isolated populations (A-B) and used to generate cDNA libraries with NEXTflex™ small RNA-Seq kit (C). Different concentrations of RPFs pooled from 2 FACS sessions (A) were separated by denaturing PAGE (B) and bands corresponding to monosome- (red box) and polysome-bound (green box) RPFs were excised and recovered. 25% of adapter concentration for 3' and 5' adapter ligations were used to prepare cDNA libraries for both monosome (red box) and polysome (green box) fractions with 25 PCR cycles (C). Due to differences in RNA concentrations, samples generated from the monosome fraction yield less cDNA than the respective polysome fraction for each germ cell type. The absence of an adapter dimer band (~130 bp) in the negative control suggests that cDNA libraries from RPFs comprise fragments between 100-200 bp. Spg: Spermatogonia; Spc I: primary Spermatocytes; rSpd: round spermatids; eSpd: elongating spermatids. Denaturing PAGE was performed with 15% TBE-Urea gels (Bio-rad) and stained with SYBR gold (Invitrogen) for visualization. A 10-bp DNA ladder (Invitrogen; Lanes 1 and 10 in B) and 26-34 nt oligomarkers (IDT; Lanes 2 and 9 in B) were used as size references. PAGE was performed with a 10% TBE gel (Bio-rad) and stained with SYBR gold (Invitrogen) for visualization. A ready to load Low MW Ladder provided with the kit was used for size reference.

fragmentation, generating footprints with variable sizes (Fig. 3A-B; Wolin and Walter (1988); Ingolia *et al.* (2009); Ingolia *et al.* (2012)). Typically, isolated monosome-bound RNA fragments are ~26-34nt and larger molecules can represent RNA protected from digestion by either more

than one ribosome or other non-ribosomal RNA binding proteins such as ribonucleoprotein (RNP) complexes (Fig. 3A-B; Ingolia *et al.* (2014)). Estimated ribosome density has been determined to be of on average 0.14 ribosomes per 100 nts for ORFs >3,600 nts, increasing to 1.2 ribosomes for ORFs <400 nts (Arava *et al.* 2003). Importantly, even with increased ribosome density, only a maximum of 5 ribosomes were found to bind to short ORFs (<400). Since the generation of footprints longer than 200 nts would require the close proximity of at least 7 individual ribosomes to prevent nuclease digestion (Wolin and Walter 1988), it seems unlikely that such fragments would correspond to polysome-bound RNA footprints. Therefore, for the sake of this analysis, we consider that RPFs <200 nts represent active translation (monosome fraction: 26-34 nts; polysome fraction: 35-200 nts), while RPFs >200 nts should result from translationally repressed mRNAs, protected from degradation, and consequently from nuclease digestion, by RNA-binding proteins (RBPs). Quantification of these RNAs provides a generalized idea of translational dynamics throughout spermatogenesis (Fig.6C). Over 700 translationally regulated transcripts have been identified, mainly by the comparison of transcriptome and proteomic data (Gan *et al.* 2013), or by array-based identification of transcripts isolated from polysome fractions (Iguchi *et al.* 2006). Our data suggests that Spg and eSpd accumulate more RNA in RNP complexes, in agreement with the peak in translational repression found for Spg to Spc transition (Gan *et al.* 2013) and the high accumulation of silenced mRNAs in the cytoplasm of spermatids (Reviewed in Kleene and Cullinane (2011)). Considering the transcription shut down during meiotic prophase (Bradley *et al.* 2004; Wang *et al.* 2005) and from mid-spermiogenesis onwards (Kierszenbaum and Tres 1978; Hecht 1998; Rolland *et al.* 2008), our observations fit in the current view that translational repression controls the timing of protein synthesis during periods of transcriptional inactivity (Kleene 2003; Idler and Yan 2012). Moreover, this accumulation of repressed transcripts is in agreement with the higher transcription/translation ratios detected for Spg and eSpd (Fig.3C). Although this ratio was perhaps expected to be lower, it is possible that the number of mouse mRNAs transcribed in eSpd and repressed for later translation, such as rat Tnp2 (Kistler *et al.* 1996), is higher than previously anticipated. Transcription rates seem to be higher in Spg, followed by spermatocytes and strikingly reduced in Spd (Fig.3C). However, given the effective proportion of each germ cell type in the seminiferous tubules (Mays-Hoopes *et al.* 1995), spermatocytes and spermatids are the most significant contributors for testis germ cell transcriptome (Soumillon *et al.* 2013). Collectively, these observations confirm that we have successfully recovered, in parallel, both total RNA and ribosome footprints from the same population of Ho-FACS sorted murine male germ cells which allows a semi-quantitative analysis.

Despite the overall high yield obtained with Ho-FACS (0.5-13x10^6 cells), the proportion of mRNA undergoing translation is low and thus minimum average amount of RPFs recovered (Fig.3C) was insufficient for the standard ribosome profiling protocol (Ingolia *et al.* 2012). Therefore we investigated the applicability of a commercial kit for library preparation of RPFs, described here in detail. Technically, the major differences of our workflow (Fig.1) to the standard ribosome profiling are: i) isolation of ribosomes by size exclusion chromatography, instead of ultracentrifugation; ii) cDNA library preparation with NEXTflex™ small RNA-Seq kit (BIOO SCIENTIFIC®) and iii) no rRNA removal step. In practice, these translate into significant reductions in costs and time, from ~3 weeks (with overnight precipitations) to ~8 days, allied to a more straightforward procedure.

A ribosome profiling dataset generated here with the standard protocol using SEC correlates with the one obtained by Castaneda *et al.* (2014) using ultracentrifugation (Supp. Fig.1). SEC has already been used for ribosome recovery (Ingolia *et al.* 2014) and our results support that SEC is a suitable, faster and less labor intensive alternative technique to recover ribosome footprints from mouse testis.

Deep-sequencing of cDNA libraries prepared from RPFs using NEXTflex™ small RNA-Seq kit (BIOO SCIENTIFIC®) shows a decrease in the percentage of reads mapping to the mouse reference genome (Supplementary Table 2) likely due to an increase of adapter dimer formation associated with smaller amounts of starting material (Fig. 4). This is a recognized technical issue of library preparation, especially for small RNAs (Kawano *et al.* 2010), which becomes critical for low input samples. PCR amplification exponentially boosts the proportion of adapter dimers and sequencing reads are wasted with useless information, decreasing the overall sample coverage. Moreover, despite the extreme efficiency in the reduction of ligation bias (Zhuang *et al.* 2012), the use of randomized adapters limits downstream processing for adapter dimer removal. Although the workflow of this kit has been optimized to avoid this issue, an appropriate adapter/RNA ratio is needed for an efficient ligation without side product formation (Song *et al.* 2014). In this case, it appears that the optimum adapter concentration for 100ng of RPFs is 25% of the stock solution, as seen by the overall increase in cDNA yield (Fig.4B), although it needs confirmation by analysis of ribo-seq data. Furthermore, this adjustment was required since our results suggested that some cDNA fragments overlapped with adapter dimers during PAGE purification of RPF libraries (Figs.4&7). Remarkably, although we skipped the rRNA depletion step to avoid further RNA loss during library preparation, we found that less than 6% of reads mapped to known rRNA sequences indicating limited contamination and coverage loss (Supplementary table 2). Preliminary ribo-seq data

generated using different methods for library preparation shows poor correlation and call for an in-depth analysis of the dataset to identify the biological relevance of the differences detected. Nonetheless, data generated using different RNA inputs suggests that with this kit the amount of starting material is not a limitation since the datasets show a good correlation in terms of captured transcripts (Fig.6).

Finally, we produced cDNA libraries for different male germ cell populations isolated by Ho-FACS (Fig.7). Given the average amount of RPFs generated for different populations (Fig.3C, Fig. 7A and Supp. Table 1), the cDNA yield detected by PAGE was expected to be higher and most likely resulted from RNA degradation during DNase treatment prior to library preparation. Additional work is required to confirm these results and evaluate the quantitative power of this technique. Specifically, the potential to detect and analyze translation rates and usage of upstream open-reading frames (uORFs) needs further investigation. Nonetheless, our approach seems to be suitable for qualitative ribosome profiling of different male germ cells.

Here, we provide the first evidence for the application of ribosome profiling to four Ho-FACS sorted male germ cell populations. We describe a faster, more affordable and practical workflow (Fig.1) that is suitable low RNA input samples, overcoming the major limitations of the standard ribosome profiling technique. This procedure can be extended to other germ cell types that can be discriminated by Ho-FACS, but importantly, it is potentially applicable to any cell population isolated from different tissues. We believe that upon further analysis and validations, our work has the potential to contribute to many fields of study, such as developmental biology, neurobiology and immunology, by bringing new tools to explore the extent and significance of cellular heterogeneity in complex tissues.

Acknowledgements

We thank the Alvin J. Siteman Cancer Center at Washington University School of Medicine and Barnes-Jewish Hospital in St. Louis, Mo., for the use of the Siteman Flow Cytometry Core, which provided staff-operated cell sorting service. The Siteman Cancer Center is supported in part by an NCI Cancer Center Support Grant #P30 CA91842. We also thank Chengran Yang and Amy Wilfert from the Washington University in St. Louis for the contribuition in the troubleshooting of the data analysis.

Competing Interests

All authors declare no competing interests.

Author's contributions

ACL obtained and processed the tissue, performed all experiments and analyzed the results. DFC and NH and analyzed the sequencing data. ACL wrote the manuscript with input from all authors. AML and DFC supervised the project.

Funding

This work was funded by an FCT doctoral fellowship [SFRH/BD/51695/2011 to ACL], grants from the United States National Institutes of Health [R01HD078641 and R01MH101810 to DFC], an FCT research contract [IF/01262/2014 to AML] and a Luso-American Development Foundation travel grant [to ACL].

References

Arava, Y., Wang, Y., Storey, J.D., Liu, C.L., Brown, P.O. and Herschlag, D. (2003). Genome-wide analysis of mRNA translation profiles in Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A,* **100**: 7, 3889-3894.

Bastos, H., Lassalle, B., Chicheportiche, A., Riou, L., Testart, J., Allemand, I. and Fouchet, P. (2005). Flow cytometric characterization of viable meiotic and postmeiotic cells by Hoechst 33342 in mouse spermatogenesis. *Cytometry A,* **65**: 1, 40-49.

Bradley, J., Baltus, A., Skaletsky, H., Royce-Tolland, M., Dewar, K. and Page, D.C. (2004). An X-to-autosome retrogene is required for spermatogenesis in mice. *Nat Genet,* **36**: 8, 872-876.

Castaneda, J., Genzor, P., van der Heijden, G.W., Sarkeshik, A., Yates, J.R., 3rd, Ingolia, N.T. and Bortvin, A. (2014). Reduced pachytene piRNAs and translation underlie spermiogenic arrest in Maelstrom mutant mice. *EMBO J,* **33**: 18, 1999-2019.

Chalmel, F., Rolland, A.D., Niederhauser-Wiederkehr, C., Chung, S.S., Demougin, P., Gattiker, A., Moore, J., Patard, J.J., Wolgemuth, D.J., Jegou, B. and Primig, M. (2007). The conserved transcriptome in human and rodent male gametogenesis. *Proc Natl Acad Sci U S A,* **104**: 20, 8346-8351.

DeLuca, D.S., Levin, J.Z., Sivachenko, A., Fennell, T., Nazaire, M.D., Williams, C., Reich, M., Winckler, W. and Getz, G. (2012). RNA-SeQC: RNA-seq metrics for quality control and process optimization. *Bioinformatics,* **28**: 11, 1530-1532.

Fallahi, M., Getun, I.V., Wu, Z.K. and Bois, P.R. (2010). A Global Expression Switch Marks Pachytene Initiation during Mouse Male Meiosis. *Genes (Basel),* **1**: 3, 469-483.

Gan, H., Cai, T., Lin, X., Wu, Y., Wang, X., Yang, F. and Han, C. (2013). Integrative proteomic and transcriptomic analyses reveal multiple post-transcriptional regulatory mechanisms of mouse spermatogenesis. *Mol Cell Proteomics,* **12**: 5, 1144-1157.

Gaysinskaya, V., Soh, I.Y., van der Heijden, G.W. and Bortvin, A. (2014). Optimized flow cytometry isolation of murine spermatocytes. *Cytometry A,* **85**: 6, 556-565.

Gerashchenko, M.V., Lobanov, A.V. and Gladyshev, V.N. (2012). Genome-wide ribosome profiling reveals complex translational regulation in response to oxidative stress. *Proc Natl Acad Sci U S A,* **109**: 43, 17394-17399.

Getun, I.V., Torres, B. and Bois, P.R. (2011). Flow cytometry purification of mouse meiotic cells. *J Vis Exp,* 50,

Hecht, N.B. (1998). Molecular mechanisms of male germ cell differentiation. *Bioessays,* **20**: 7, 555-561.

Idler, R.K. and Yan, W. (2012). Control of messenger RNA fate by RNA-binding proteins: an emphasis on mammalian spermatogenesis. *J Androl,* **33**: 3, 309-337.

Iguchi, N., Tobias, J.W. and Hecht, N.B. (2006). Expression profiling reveals meiotic male germ cell mRNAs that are translationally up- and down-regulated. *Proc Natl Acad Sci U S A,* **103**: 20, 7712-7717.

Ingolia, N.T., Brar, G.A., Rouskin, S., McGeachy, A.M. and Weissman, J.S. (2012). The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. *Nat Protoc,* **7**: 8, 1534-1550.

Ingolia, N.T., Brar, G.A., Stern-Ginossar, N., Harris, M.S., Talhouarne, G.J., Jackson, S.E., Wills, M.R. and Weissman, J.S. (2014). Ribosome profiling reveals pervasive translation outside of annotated protein-coding genes. *Cell Rep,* **8**: 5, 1365-1379.

Ingolia, N.T., Ghaemmaghami, S., Newman, J.R. and Weissman, J.S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science,* **324**: 5924, 218-223.

Ingolia, N.T., Lareau, L.F. and Weissman, J.S. (2011). Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell,* **147**: 4, 789-802.

Jackson, R. and Standart, N. (2015). The awesome power of ribosome profiling. *RNA,* **21**: 4, 652-654.

Kawano, M., Kawazu, C., Lizio, M., Kawaji, H., Carninci, P., Suzuki, H. and Hayashizaki, Y. (2010). Reduction of non-insert sequence reads by dimer eliminator LNA oligonucleotide for small RNA deep sequencing. *Biotechniques,* **49**: 4, 751-755.

Kierszenbaum, A.L. and Tres, L.L. (1978). RNA transcription and chromatin structure during meiotic and postmeiotic stages of spermatogenesis. *Fed Proc,* **37**: 11, 2512-2516.

Kistler, W.S., Henriksen, K., Mali, P. and Parvinen, M. (1996). Sequential expression of nucleoproteins during rat spermiogenesis. *Exp Cell Res,* **225**: 2, 374-381.

Kleene, K.C. (2003). Patterns, mechanisms, and functions of translation regulation in mammalian spermatogenic cells. *Cytogenet Genome Res,* **103**: 3-4, 217-224.

Kleene, K.C. and Cullinane, D.L. (2011). Maybe repressed mRNAs are not stored in the chromatoid body in mammalian spermatids. *Reproduction,* **142**: 3, 383-388.

Larsson, O., Tian, B. and Sonenberg, N. (2013). Toward a genome-wide landscape of translational control. *Cold Spring Harb Perspect Biol,* **5**: 1, a012302.

Lee, S., Liu, B., Lee, S., Huang, S.X., Shen, B. and Qian, S.B. (2012). Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. *Proc Natl Acad Sci U S A,* **109**: 37, E2424-2432.

Li, H. and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics,* **25**: 14, 1754-1760.

Mays-Hoopes, L.L., Bolen, J., Riggs, A.D. and Singer-Sam, J. (1995). Preparation of spermatogonia, spermatocytes, and round spermatids for analysis of gene expression using fluorescence-activated cell sorting. *Biol Reprod,* **53**: 5, 1003-1011.

Piccirillo, C.A., Bjur, E., Topisirovic, I., Sonenberg, N. and Larsson, O. (2014). Translational control of immune responses: from transcripts to translatomes. *Nat Immunol,* **15**: 6, 503-511.

R Development Core Team (2012). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*,

Rolland, A.D., Jegou, B. and Pineau, C. (2008). Testicular development and spermatogenesis: harvesting the postgenomics bounty. *Adv Exp Med Biol,* **636**: 16-41.

Song, Y., Liu, K.J. and Wang, T.H. (2014). Elimination of ligation dependent artifacts in T4 RNA ligase to achieve high efficiency and low bias microRNA capture. *PLoS One,* **9**: 4, e94619.

Soumillon, M., Necsulea, A., Weier, M., Brawand, D., Zhang, X., Gu, H., Barthes, P., Kokkinaki, M., Nef, S., Gnirke, A., Dym, M., de Massy, B., Mikkelsen, T.S. and Kaessmann, H. (2013). Cellular source and mechanisms of high transcriptome complexity in the mammalian testis. *Cell Rep,* **3**: 6, 2179-2190.

Vogel, C. and Marcotte, E.M. (2012). Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat Rev Genet,* **13**: 4, 227-232.

Wang, P.J., Page, D.C. and McCarrey, J.R. (2005). Differential expression of sex-linked and autosomal germ-cell-specific genes during spermatogenesis in the mouse. *Hum Mol Genet,* **14**: 19, 2911-2918.

Wolin, S.L. and Walter, P. (1988). Ribosome pausing and stacking during translation of a eukaryotic mRNA. *EMBO J,* **7**: 11, 3559-3569.

Zhuang, F., Fuchs, R.T., Sun, Z., Zheng, Y. and Robb, G.B. (2012). Structural bias in T4 RNA ligase-mediated 3'-adapter ligation. *Nucleic Acids Res,* **40**: 7, e54.

Discussion

Over the past decade, the technological advances have allowed for a systematic assessment of genetic variation, mRNA expression and protein levels in different biological systems in large scales. With the use of array- and sequencing-based techniques to identify genetic causes of male infertility in patient cohorts, it has become increasingly evident that the genetic component of this disorder is extremely complex and that attributing causality to the variants showing association to disease remains a challenge (MacArthur *et al.* 2014; Carrell *et al.* 2016). Indeed, the work described in Chapter 1 emphasizes some of the issues revealed by candidate-gene approaches when addressing the genetic basis of male infertility. Therefore, in the second part of my project, I sought to investigate the basic biology of spermatogenesis to provide support to a better understanding of this disease. The rational was based on a few core observations/ideas:

- ✓ The biological mechanisms driving spermatogenesis progression are still poorly understood;
- ✓ Combined effects of mild perturbations to any of the regulatory layers (from gene to protein) controlling the normal course of gamete development can potentially unbalance the system and lead to disease;
- ✓ The identification of molecules involved in this process provides an invaluably comprehensive catalogue of functionally relevant genes, potentially altered in infertile male patients;
- ✓ Comparative studies of normal spermatogenesis progression in different species highlight both the core elements of this process, conserved throughout evolution, as well as the unique fine-tuning mechanisms that control species-specific attributes of gametes;
- ✓ Post-transcriptional and translational regulation of mRNAs play a key role in the maintenance of normal gamete formation and are the least elucidated regulatory mechanisms.

Considering the above, and as previously detailed, the main goal of this part of my work was to develop and apply a protocol to isolate mammalian male germ cells of different species for high-throughput studies that interrogate the mechanisms involved in spermatogenesis progression, such as translational regulation, with an evolutionary approach. These can

contribute to a better understanding of the underlying biology of spermatogenesis and allow the assessment of the functional impact of variation found in the context of male infertility.

Ho-FACS as a transversal method for isolation of male germ cells from different mammalian species

The cellular heterogeneity of the testis, and resulting confounding factors when performing molecular analyses in the whole testicular tissue, call for efficient techniques for the isolation of germ cells in distinct developmental stages. Ho-FACS of murine germ cells has been extensively optimized over the past decade and currently allows for the isolation of 9 cell types in distinct developmental stages (Bastos *et al.* 2005; Geisinger and Rodriguez-Casuriaga 2010; Getun *et al.* 2011; Gaysinskaya *et al.* 2014; Gaysinskaya and Bortvin 2015). This technique is based on measurements of chromatin amount and structure detectable using the fluorescent Hoechst-33342 DNA dye. Flow cytometry of testicular cell suspensions from non-mouse mammalian species using different dyes, staining protocols and flow cytometry parameters for analysis have revealed similar profiles in terms of DNA ploidy/stainability (Reviewed in Geisinger and Rodriguez-Casuriaga (2010)). In paper IV I investigated the applicability of Ho-FACS as a transversal method for isolation of male germ cells from different mammalian species. Although testis dissociation protocols needed species-specific adjustments, this work provides the first proof of principle that a standardized Ho-FACS protocol can indeed be used in other mammalian species such as dog and rat (Paper IV- Fig. 3). One way to circumvent this issue, is to apply mechanical rather than enzymatic tissue dissociation. This approach was already described for rat, mouse and guinea pig testis dissociation and shown to yield good quality cell suspensions that could be used for flow cytometry (Rodriguez-Casuriaga *et al.* 2013).

When applying genome technology to the study of spermatogenesis, the yield and purity of isolated cell types are of the outmost importance for the success and accuracy of the study. Although lower than that of Ho-FACS sorted cells from the mouse, a fairly high 80% average purity was estimated for cell populations obtained for dog and rat, indicating the successful application of the Ho-FACS protocol to other non-mouse mammalian species (Paper IV- Fig. 4). Moreover, FACS processing for only 1.5h yielded sufficient material for further downstream high-throughput studies with an average of 6.0 X 10^6 cells recovered. Processing time is a key factor for such experiments given the need to preserve the integrity of cellular components of

interest such as mRNAs. Further work involves using blue Ho fluorescence to gate germ cells based on ploidy (1N, 2N and 4N) as previously applied in other studies (Gaysinskaya *et al.* 2014; Gaysinskaya and Bortvin 2015). This strategy should decrease the level of contamination of populations with other cell-types that could further be quantified by RT-PCR of germ cell type specific transcripts or global RNA-seq analysis. Measuring purity of cell populations based on gene expression also overcomes the inherent subjectivity associated with morphological evaluation of sorted cells using microscopy, despite the latter being recurrently used to assess the purity of cell populations. Moreover, to take full advantage of the technique, the possibility to isolate other germ cell types other than Spg, Spc I and Spd, needs to be investigated. Considering previous works, these include further resolving the Spc I population into 4 subpopulations in different stages of meiotic prophase and the Spd population in round and elongating. Indeed, the results shown in paper IV (Fig. 6) for the separation of round and elongating Spd populations based on size and cellular complexity (FS and SS) suggest that the same strategy should, in theory, be applicable to other mammalian species.

Concerns related to cellular toxicity induced by UV excitation of Ho led to the design and investigation of new fluorescent dyes for flow cytometry of testicular cell suspensions. While some have exhibited adverse genotoxic effects, the SYTO® fluorochromes (Molecular Probes/Invitrogen) show promising results (Geisinger and Rodriguez-Casuriaga 2010). In fact SYTO16 was recently used to discriminate 4 subpopulations of mouse spermatids (Simard *et al.* 2014; Simard *et al.* 2015), but its applicability to discern germ cells throughout spermatogenesis remains unexplored. Another supravital dye, Vybrant® DyeCycle™ Green Stain (VDG; Molecular Probes/Invitrogen), was shown to enable the isolation of high-purity meiotic substages from the mouse and guinea pig testes (Rodriguez-Casuriaga *et al.* 2014). However, the FACS plots obtained using VDG staining appear less intuitive and straightforward to define gates when compared to Ho blue/red fluorescence plots. Although Ho has been shown to induce variable DNA damage responses dependent on the duration of treatment and dye concentration (Zhao *et al.* 2009), transcriptome data for mouse Spg, early/late Spc I, Spc II and Spd generated at the Conrad lab using Ho-FACS (data not shown) correlates with previous studies using other methods for germ cell isolation. Comparison of gene expression data from FACS sorted cells using the new non-toxic dyes or Ho should provide insights on how different staining methods affect downstream analysis of molecular profiles.

A recent development in flow cytometry techniques brought the exciting possibility of cell separation based on visual cellular traits. This new technology, called imaging flow

cytometry (IFC), combines features of flow cytometry, fluorescent microscopy and data-processing algorithms to process thousands of cellular events using a multiparametric fluorescent and morphological analysis [Reviewed in Barteneva *et al.* (2012)]. Combining the Ho-FACS technology with IFC could allow the identification and isolation of male germ cells based on real images of chromatin distribution and condensation, which exhibit unique and well characterized patterns depending on the developmental stage of the cell. IFC is a major breakthrough holding the promise to open new avenues of research in male reproductive biology and diagnostics of male infertility by allowing the simultaneous analysis of morphological features and visualization of cellular dynamics of specific markers.

A new strategy for translational profiling of murine male germ cells isolated by Ho-FACS

Several studies have demonstrated the importance of post-transcriptional and translation control of mRNAs for normal spermatogenesis progression (See Bettegowda and Wilkinson (2010); Idler and Yan (2012); de Mateo and Sassone-Corsi (2014); Yadav and Kotaja (2014)). These mechanisms are significantly more important in the testes when compared to other tissues/cell types and have presumably evolved as a compensation for the major transcription repression occurring during meiotic prophase (Bradley *et al.* 2004; Wang *et al.* 2005) and from mid-spermiogenesis onwards (Kierszenbaum and Tres 1978; Hecht 1998; Rolland *et al.* 2008). Therefore, addressing translational control throughout spermatogenesis not only reveals the intricate network of germ cell regulation but is also an excellent model to explore the basic mechanisms involved in this type of molecular control. Making use of the optimized Ho-FACS technique (Paper IV), I adopted a new strategy for deep-sequencing of mRNA molecules undergoing translation in different male germ cells (Paper V- Fig. 1). These modifications were required since the amount of ribosome-protected fragments recovered from germs cells are below the minimum of 2 µg used in the original protocol (500-800 ng; Paper V- Fig. 6). In fact, the amount of starting material is the most limiting factor for most high-throughput studies, including the performance of massive next-generation sequencing. Another major concern in such studies is reproducibility, since working with low RNA yields introduces higher experimental biases occurring mainly during the amplification steps. I reasoned that library preparation with a standardized commercial library prep kit, optimized for low input samples, would be the best approach to ensure reproducibility while reducing experimentally introduced biases. Also, recent reports have pointed to the issue of preferential adaptor ligation

during library preparation, which can result in underrepresentation or failure to detect RNA molecules containing specific 3' sequences (Zhuang *et al.* 2012). Keeping this in mind, I selected the NEXTflex™ small RNA-Seq kit (BIOO SCIENTIFIC®), optimized for preparation of miRNA cDNAs from 1-10 µg of total RNA. Also, this is the only product that utilizes adapters with randomized ends to greatly reduce sequence bias. With this approach, I was able to prepare cDNA libraries of mRNAs undergoing translation in 4 different germ cell types: Spg, Spc I, rSpd and eSpd (Paper V- Fig. 6). Although library preparation with this kit includes several steps of adaptor depletion with magnetic beads and a special depletion solution, the major source of technical issues with this protocol is still the formation of adapter dimers during PCR amplification (Paper V- Figs. 3 & 6). Moreover, the results suggest that a fraction of cDNA molecules might present a similar migration pattern as the adapter dimers during PAGE purification of the libraries (Paper V- Figs. 3 & 6). Similarly to what has been described for the generation of miRNA libraries (Song *et al.* 2014), it seems that a proper adapter/RNA ratio is crucial to avoid the excess of 3' and 5' adapters while maintaining ligation efficiency. This has proven to be a very challenging task and a few other conditions (e.g. removing DNase treatment and performing overnight ligation reactions) should be tested to improve the current efficiency of this protocol. Nevertheless, I show that translational profiling of murine male germ cells is achievable, at least in a qualitative matter. Pooling together more biological samples isolated in independent Ho-FACS experiments can potentially be used as a strategy to increase the cell number and therefore the amount of starting material for library preparation. This should decrease the experimental noise in the sequencing data and allow for a quantitative analysis of translation rates of different transcripts during spermatogenesis.

Upon validation of the quantitative power of this optimized protocol, extending it to other germ cell types that can be isolated by Ho-FACS can answer many outstanding biological questions about translation regulation in spermatogenesis. First, it allows for in-depth genome-wide measurements of germ cell type-specific mRNA translation rates as well as programmed translation of non-canonical isoforms and non-coding RNAs. Since this workflow was designed to include parallel sequencing of both total RNA and RPFs, mRNA transport/translation and translational efficiency of different transcripts as well as of differential translation (Ingolia *et al.* 2009; Ingolia *et al.* 2012; Larsson *et al.* 2013) can be measured and identified. Moreover, if we assume that the RNA molecules longer than 200 nt (RPFs >200) isolated during size exclusion chromatography represent transcripts stored in RNPs, as discussed in paper V, this technique has the potential to register the dynamics of translational repression and identify key molecules undergoing such process. However, the nature of the transcripts present in the fraction needs

further validation with adequate techniques. One way to do this would be to target RNAs sequenced from the RPFs>200 fraction with RNA-affinity purification methods followed by mass spectrometry analysis to identify the RBPs (Reviewed in Faoro and Ataide (2014)) bound to those RNAs. Quantification of alternative translation initiation sites (TIS), upstream open reading frames (uORFs) and the kinetics of translation can also be achieved by treating cells with Lactomidomycin, which stops translation specifically at the TIS, prior to ribosome profiling (Ingolia *et al.* 2011; Ingolia *et al.* 2012; S. Lee *et al.* 2012). Comparing data obtained for different germ cells and/or whole testis with other tissues could potentially reveal new mechanisms of translational control and highlight its importance for proper gamete formation. Also, the identification of uORFs can expose previously unknown functionally relevant regions in which genetic variation could have deleterious consequences. All these data can then be collectively compared for different germ cell types and potentially shed some light into the importance of specific mechanisms and transcripts in particular developmental stages. This would contribute to the functional annotation of fertility relevant genes and potentially help establish genotype/phenotype correlations in infertile men.

Overall, the approach described in paper V (Fig. 1) for ribosome profiling of murine male germ cells represents a faster, more affordable and practical workflow that is suitable low RNA input samples, overcoming the major limitations of the standard ribosome profiling technique. Recent works have already reported major improvements on the original protocol (Aeschimann *et al.* 2015; Chung *et al.* 2015; Miettinen and Bjorklund 2015; Reid *et al.* 2015), however none has yet focused on decreasing the minimum amount of starting material. The work reported here brings for the first time the possibility of studying translational regulation in male germ cells of different developmental stages. Importantly, this protocol can be applied to other challenging models, such as the brain and the immune system, where cellular heterogeneity prevents straightforward and informative high-throughput studies. Moreover, combining the tools described in papers IV and V will allow for comparative studies of translational dynamics throughout spermatogenesis. This approach can bring fundamental knowledge about the evolution of the basic process of translation as well as translation regulation in different mammalian clades.

Literature Cited

Aeschimann, F., Xiong, J., Arnold, A., Dieterich, C. and Grosshans, H. (2015). Transcriptome-wide measurement of ribosomal occupancy by ribosome profiling. *Methods,* **85**: 75-89.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H.*, et al.* (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet,* **25**: 1, 25-29.

Barteneva, N.S., Fasler-Kan, E. and Vorobjev, I.A. (2012). Imaging flow cytometry: coping with heterogeneity in biological systems. *J Histochem Cytochem,* **60**: 10, 723-733.

Bastos, H., Lassalle, B., Chicheportiche, A., Riou, L., Testart, J., Allemand, I. and Fouchet, P. (2005). Flow cytometric characterization of viable meiotic and postmeiotic cells by Hoechst 33342 in mouse spermatogenesis. *Cytometry A,* **65**: 1, 40-49.

Belloc, S., Benkhalifa, M., Cohen-Bacrie, M., Dalleac, A., Chahine, H., Amar, E. and Zini, A. (2014). Which isolated sperm abnormality is most related to sperm DNA damage in men presenting for infertility evaluation. *J Assist Reprod Genet,* **31**: 5, 527-532.

Bendall, S.C., Nolan, G.P., Roederer, M. and Chattopadhyay, P.K. (2012). A deep profiler's guide to cytometry. *Trends Immunol,* **33**: 7, 323-332.

Bettegowda, A. and Wilkinson, M.F. (2010). Transcription and post-transcriptional regulation of spermatogenesis. *Philos Trans R Soc Lond B Biol Sci,* **365**: 1546, 1637-1651.

Bonner, W.A., Hulett, H.R., Sweet, R.G. and Herzenberg, L.A. (1972). Fluorescence activated cell sorting. *Rev Sci Instrum,* **43**: 3, 404-409.

Bradley, J., Baltus, A., Skaletsky, H., Royce-Tolland, M., Dewar, K. and Page, D.C. (2004). An X-to-autosome retrogene is required for spermatogenesis in mice. *Nat Genet,* **36**: 8, 872-876.

Bryant, J.M., Meyer-Ficca, M.L., Dang, V.M., Berger, S.L. and Meyer, R.G. (2013). Separation of spermatogenic cell types using STA-PUT velocity sedimentation. *J Vis Exp,* 80,

Cagney, G., Park, S., Chung, C., Tong, B., O'Dushlaine, C., Shields, D.C. and Emili, A. (2005). Human tissue profiling with multidimensional protein identification technology. *J Proteome Res,* **4**: 5, 1757-1767.

Calvel, P., Rolland, A.D., Jegou, B. and Pineau, C. (2010). Testicular postgenomics: targeting the regulation of spermatogenesis. *Philos Trans R Soc Lond B Biol Sci,* **365**: 1546, 1481-1500.

Carrell, D.T., Aston, K.I., Oliva, R., Emery, B.R. and De Jonge, C.J. (2016). The "omics" of human male infertility: integrating big data in a systems biology approach. *Cell Tissue Res,* **363**: 1, 295-312.

Castaneda, J., Genzor, P., van der Heijden, G.W., Sarkeshik, A., Yates, J.R., 3rd, Ingolia, N.T. and Bortvin, A. (2014). Reduced pachytene piRNAs and translation underlie spermiogenic arrest in Maelstrom mutant mice. *EMBO J,* **33**: 18, 1999-2019.

Chalmel, F., Com, E., Lavigne, R., Hernio, N., Teixeira-Gomes, A.P., Dacheux, J.L. and Pineau, C. (2014). An integrative omics strategy to assess the germ cell secretome and to decipher sertoli-germ cell crosstalk in the Mammalian testis. *PLoS One,* **9**: 8, e104418.

Chalmel, F., Lardenois, A. and Primig, M. (2007a). Toward understanding the core meiotic transcriptome in mammals and its implications for somatic cancer. *Ann N Y Acad Sci,* **1120**: 1-15.

Chalmel, F., Rolland, A.D., Niederhauser-Wiederkehr, C., Chung, S.S., Demougin, P., Gattiker, A., Moore, J., Patard, J.J., Wolgemuth, D.J., Jegou, B. and Primig, M. (2007b). The conserved transcriptome in human and rodent male gametogenesis. *Proc Natl Acad Sci U S A,* **104**: 20, 8346-8351.

Chang, Y.F., Lee-Chang, J.S., Panneerdoss, S., MacLean, J.A., 2nd and Rao, M.K. (2011). Isolation of Sertoli, Leydig, and spermatogenic cells from the mouse testis. *Biotechniques,* **51**: 5, 341-342, 344.

Chattopadhyay, P.K. and Roederer, M. (2012). Cytometry: today's technology and tomorrow's horizons. *Methods,* **57**: 3, 251-258.

Chen, G., Gharib, T.G., Huang, C.C., Taylor, J.M., Misek, D.E., Kardia, S.L., Giordano, T.J., Iannettoni, M.D., Orringer, M.B., Hanash, S.M. and Beer, D.G. (2002). Discordant protein and mRNA expression in lung adenocarcinomas. *Mol Cell Proteomics,* **1**: 4, 304-313.

Chen, T., Zhao, J., Ma, J. and Zhu, Y. (2015). Web resources for mass spectrometry-based proteomics. *Genomics Proteomics Bioinformatics,* **13**: 1, 36-39.

Chocu, S., Calvel, P., Rolland, A.D. and Pineau, C. (2012). Spermatogenesis in mammals: proteomic insights. *Syst Biol Reprod Med,* **58**: 4, 179-190.

Chocu, S., Evrard, B., Lavigne, R., Rolland, A.D., Aubry, F., Jegou, B., Chalmel, F. and Pineau, C. (2014). Forty-four novel protein-coding loci discovered using a proteomics informed by transcriptomics (PIT) approach in rat male germ cells. *Biol Reprod,* **91**: 5, 123.

Chung, B.Y., Hardcastle, T.J., Jones, J.D., Irigoyen, N., Firth, A.E., Baulcombe, D.C. and Brierley, I. (2015). The use of duplex-specific nuclease in ribosome profiling and a user-friendly software package for Ribo-seq data analysis. *RNA,* **21**: 10, 1731-1745.

Com, E., Melaine, N., Chalmel, F. and Pineau, C. (2014). Proteomics and integrative genomics for unraveling the mysteries of spermatogenesis: the strategies of a team. *J Proteomics,* **107**: 128-143.

Conrads, K.A., Yi, M., Simpson, K.A., Lucas, D.A., Camalier, C.E., Yu, L.R., Veenstra, T.D., Stephens, R.M., Conrads, T.P. and Beck, G.R., Jr. (2005). A combined proteome and microarray investigation of inorganic phosphate-induced pre-osteoblast cells. *Mol Cell Proteomics,* **4**: 9, 1284-1296.

Cordelli, E., Eleuteri, P., Leter, G., Rescia, M. and Spano, M. (2005). Flow cytometry applications in the evaluation of sperm quality: semen analysis, sperm function and DNA integrity. *Contraception,* **72**: 4, 273-279.

De Iuliis, G.N., Thomson, L.K., Mitchell, L.A., Finnie, J.M., Koppers, A.J., Hedges, A., Nixon, B. and Aitken, R.J. (2009). DNA damage in human spermatozoa is highly correlated with the efficiency of chromatin remodeling and the formation of 8-hydroxy-2'-deoxyguanosine, a marker of oxidative stress. *Biol Reprod,* **81**: 3, 517-524.

de Mateo, S. and Sassone-Corsi, P. (2014). Regulation of spermatogenesis by small non-coding RNAs: role of the germ granule. *Semin Cell Dev Biol,* **29**: 84-92.

Druart, X., Gatti, J.L., Huet, S., Dacheux, J.L. and Humblot, P. (2009). Hypotonic resistance of boar spermatozoa: sperm subpopulations and relationship with epididymal maturation and fertility. *Reproduction,* **137**: 2, 205-213.

Evenson, D., Darzynkiewicz, Z., Jost, L., Janca, F. and Ballachey, B. (1986). Changes in accessibility of DNA to various fluorochromes during spermatogenesis. *Cytometry,* **7**: 1, 45-53.

Evenson, D.P. and Melamed, M.R. (1983). Rapid analysis of normal and abnormal cell types in human semen and testis biopsies by flow cytometry. *J Histochem Cytochem,* **31**: 1A Suppl, 248-253.

Falciatori, I., Borsellino, G., Haliassos, N., Boitani, C., Corallini, S., Battistini, L., Bernardi, G., Stefanini, M. and Vicini, E. (2004). Identification and enrichment of spermatogonial stem cells displaying side-population phenotype in immature mouse testis. *FASEB J,* **18**: 2, 376-378.

Fang, F.C. and Casadevall, A. (2011). Reductionistic and holistic science. *Infect Immun,* **79**: 4, 1401-1404.

Faoro, C. and Ataide, S.F. (2014). Ribonomic approaches to study the RNA-binding proteome. *FEBS Lett,* **588**: 20, 3649-3664.

Faye, M.D., Graber, T.E. and Holcik, M. (2014). Assessment of selective mRNA translation in mammalian cells by polysome profiling. *J Vis Exp,* 92, e52295.

Galbraith, D. (2012). Flow cytometry and cell sorting: the next generation. *Methods,* **57**: 3, 249-250.

Gan, H., Cai, T., Lin, X., Wu, Y., Wang, X., Yang, F. and Han, C. (2013). Integrative proteomic and transcriptomic analyses reveal multiple post-transcriptional regulatory mechanisms of mouse spermatogenesis. *Mol Cell Proteomics,* **12**: 5, 1144-1157.

Garner, D.L. (2006). Flow cytometric sexing of mammalian sperm. *Theriogenology,* **65**: 5, 943-957.

Gassei, K., Ehmcke, J., Dhir, R. and Schlatt, S. (2010). Magnetic activated cell sorting allows isolation of spermatogonia from adult primate testes and reveals distinct GFRa1-positive subpopulations in men. *J Med Primatol,* **39**: 2, 83-91.

Gaysinskaya, V. and Bortvin, A. (2015). Flow cytometry of murine spermatocytes. *Curr Protoc Cytom,* **72**: 7 44 41-47 44 24.

Gaysinskaya, V., Soh, I.Y., van der Heijden, G.W. and Bortvin, A. (2014). Optimized flow cytometry isolation of murine spermatocytes. *Cytometry A,* **85**: 6, 556-565.

Geisinger, A. and Rodriguez-Casuriaga, R. (2010). Flow cytometry for gene expression studies in Mammalian spermatogenesis. *Cytogenet Genome Res,* **128**: 1-3, 46-56.

Gerashchenko, M.V., Lobanov, A.V. and Gladyshev, V.N. (2012). Genome-wide ribosome profiling reveals complex translational regulation in response to oxidative stress. *Proc Natl Acad Sci U S A,* **109**: 43, 17394-17399.

Getun, I.V., Torres, B. and Bois, P.R. (2011). Flow cytometry purification of mouse meiotic cells. *J Vis Exp,* 50,

Goodell, M.A., Brose, K., Paradis, G., Conner, A.S. and Mulligan, R.C. (1996). Isolation and functional properties of murine hematopoietic stem cells that are replicating in vivo. *J Exp Med,* **183**: 4, 1797-1806.

Govin, J., Gaucher, J., Ferro, M., Debernardi, A., Garin, J., Khochbin, S. and Rousseaux, S. (2012). Proteomic strategy for the identification of critical actors in reorganization of the post-meiotic male genome. *Mol Hum Reprod,* **18**: 1, 1-13.

Grogan, W.M., Farnham, W.F. and Sabau, J.M. (1981). DNA analysis and sorting of viable mouse testis cells. *J Histochem Cytochem,* **29**: 6, 738-746.

Grunewald, S. and Paasch, U. (2013). Sperm selection for ICSI using annexin V. *Methods Mol Biol,* **927**: 257-262.

Gygi, S.P., Rochon, Y., Franza, B.R. and Aebersold, R. (1999). Correlation between protein and mRNA abundance in yeast. *Mol Cell Biol,* **19**: 3, 1720-1730.

Haider, S. and Pal, R. (2013). Integrated analysis of transcriptomic and proteomic data. *Curr Genomics,* **14**: 2, 91-110.

He, Z., Kokkinaki, M., Jiang, J., Zeng, W., Dobrinski, I. and Dym, M. (2012). Isolation of human male germ-line stem cells using enzymatic digestion and magnetic-activated cell sorting. *Methods Mol Biol,* **825**: 45-57.

Hecht, N.B. (1998). Molecular mechanisms of male germ cell differentiation. *Bioessays,* **20**: 7, 555-561.

Hermann, B.P., Sukhwani, M., Simorangkir, D.R., Chu, T., Plant, T.M. and Orwig, K.E. (2009). Molecular dissection of the male germ cell lineage identifies putative spermatogonial stem cells in rhesus macaques. *Hum Reprod,* **24**: 7, 1704-1716.

Herzenberg, L.A., Parks, D., Sahaf, B., Perez, O., Roederer, M. and Herzenberg, L.A. (2002). The history and future of the fluorescence activated cell sorter and flow cytometry: a view from Stanford. *Clin Chem,* **48**: 10, 1819-1827.

Herzenberg, L.A., Sweet, R.G. and Herzenberg, L.A. (1976). Fluorescence-activated cell sorting. *Sci Am,* **234**: 3, 108-117.

Idler, R.K. and Yan, W. (2012). Control of messenger RNA fate by RNA-binding proteins: an emphasis on mammalian spermatogenesis. *J Androl,* **33**: 3, 309-337.

Iguchi, N., Tobias, J.W. and Hecht, N.B. (2006). Expression profiling reveals meiotic male germ cell mRNAs that are translationally up- and down-regulated. *Proc Natl Acad Sci U S A,* **103**: 20, 7712-7717.

Ingolia, N.T., Brar, G.A., Rouskin, S., McGeachy, A.M. and Weissman, J.S. (2012). The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. *Nat Protoc,* **7**: 8, 1534-1550.

Ingolia, N.T., Brar, G.A., Stern-Ginossar, N., Harris, M.S., Talhouarne, G.J., Jackson, S.E., Wills, M.R. and Weissman, J.S. (2014). Ribosome profiling reveals pervasive translation outside of annotated protein-coding genes. *Cell Rep,* **8**: 5, 1365-1379.

Ingolia, N.T., Ghaemmaghami, S., Newman, J.R. and Weissman, J.S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science,* **324**: 5924, 218-223.

Ingolia, N.T., Lareau, L.F. and Weissman, J.S. (2011). Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell,* **147**: 4, 789-802.

Jackson, R. and Standart, N. (2015). The awesome power of ribosome profiling. *RNA,* **21**: 4, 652-654.

Jenkins, J.A., Draugelis-Dale, R.O., Pinkney, A.E., Iwanowicz, L.R. and Blazer, V.S. (2015). Flow cytometric method for measuring chromatin fragmentation in fixed sperm from yellow perch (Perca flavescens). *Theriogenology,* **83**: 5, 920-931.

Kierszenbaum, A.L. and Tres, L.L. (1978). RNA transcription and chromatin structure during meiotic and postmeiotic stages of spermatogenesis. *Fed Proc,* **37**: 11, 2512-2516.

Kim, Y.H., Kim, B.J., Kim, B.G., Lee, Y.A., Kim, K.J., Chung, H.J., Hwang, S., Woo, J.S., Park, J.K., Schmidt, J.A., Pang, M.G. and Ryu, B.Y. (2013). Stage-specific embryonic antigen-1 expression by undifferentiated spermatogonia in the prepubertal boar testis. *J Anim Sci,* **91**: 7, 3143-3154.

Kleene, K.C., Distel, R.J. and Hecht, N.B. (1984). Translational regulation and deadenylation of a protamine mRNA during spermiogenesis in the mouse. *Dev Biol,* **105**: 1, 71-79.

Kordan, W., Fraser, L., Wysocki, P., Strzezek, R., Lecewicz, M., Mogielnicka-Brzozowska, M., Dziekonska, A., Soliwoda, D. and Koziorowska-Gilun, M. (2013). Semen quality assessments and their significance in reproductive technology. *Pol J Vet Sci,* **16**: 4, 823-833.

Krishnamurthy, H., Weinbauer, G.F., Aslam, H., Yeung, C.H. and Nieschlag, E. (1998). Quantification of apoptotic testicular germ cells in normal and methoxyacetic acid-treated mice as determined by flow cytometry. *J Androl,* **19**: 6, 710-717.

Lam, D.M., Furrer, R. and Bruce, W.R. (1970). The separation, physical characterization, and differentiation kinetics of spermatogonial cells of the mouse. *Proc Natl Acad Sci U S A,* **65**: 1, 192-199.

Lardenois, A., Gattiker, A., Collin, O., Chalmel, F. and Primig, M. (2010). GermOnline 4.0 is a genomics gateway for germline development, meiosis and the mitotic cell cycle. *Database (Oxford),* **2010**: baq030.

Larsson, O., Tian, B. and Sonenberg, N. (2013). Toward a genome-wide landscape of translational control. *Cold Spring Harb Perspect Biol,* **5**: 1, a012302.

Lassalle, B., Bastos, H., Louis, J.P., Riou, L., Testart, J., Dutrillaux, B., Fouchet, P. and Allemand, I. (2004). 'Side Population' cells in adult mouse testis express Bcrp1 gene and are enriched in spermatogonia and germinal stem cells. *Development,* **131**: 2, 479-487.

Lee, K., Haugen, H.S., Clegg, C.H. and Braun, R.E. (1995). Premature translation of protamine 1 mRNA causes precocious nuclear condensation and arrests spermatid differentiation in mice. *Proc Natl Acad Sci U S A,* **92**: 26, 12451-12455.

Lee, S., Liu, B., Lee, S., Huang, S.X., Shen, B. and Qian, S.B. (2012). Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. *Proc Natl Acad Sci U S A,* **109**: 37, E2424-2432.

MacArthur, D.G., Manolio, T.A., Dimmock, D.P., Rehm, H.L., Shendure, J., *et al.* (2014). Guidelines for investigating causality of sequence variants in human disease. *Nature,* **508**: 7497, 469-476.

Margolin, G., Khil, P.P., Kim, J., Bellani, M.A. and Camerini-Otero, R.D. (2014). Integrated transcriptome analysis of mouse spermatogenesis. *BMC Genomics,* **15**: 39.

Mays-Hoopes, L.L., Bolen, J., Riggs, A.D. and Singer-Sam, J. (1995). Preparation of spermatogonia, spermatocytes, and round spermatids for analysis of gene expression using fluorescence-activated cell sorting. *Biol Reprod,* **53**: 5, 1003-1011.

Meistrich, M.L. (1977). Separation of spermatogenic cells and nuclei from rodent testes. *Methods Cell Biol,* **15**: 15-54.

Michel, A.M., Mullan, J.P., Velayudhan, V., O'Connor, P.B., Donohue, C.A. and Baranov, P.V. (2016). RiboGalaxy: A browser based platform for the alignment, analysis and visualization of ribosome profiling data. *RNA Biol,* **13**: 3, 316-319.

Miettinen, T.P. and Bjorklund, M. (2015). Modified ribosome profiling reveals high abundance of ribosome protected mRNA fragments derived from 3' untranslated regions. *Nucleic Acids Res,* **43**: 2, 1019-1034.

Mooney, M. and McWeeney, S. (2014). Data integration and reproducibility for high-throughput transcriptomics. *Int Rev Neurobiol,* **116**: 55-71.

Omran, H.M., Bakhiet, M. and Dashti, M.G. (2013). DNA integrity is a critical molecular indicator for the assessment of male infertility. *Mol Med Rep,* **7**: 5, 1631-1635.

Ormerod, M.G. (2008). Flow Cytometry - A Basic Introduction. 978-0-9559812-0-3: 978-0-9559812-0-3. http://flowbook.denovosoftware.com/

Payne, S.H. (2015). The utility of protein and mRNA correlation. *Trends Biochem Sci,* **40**: 1, 1-3.

Perez-Riverol, Y., Alpi, E., Wang, R., Hermjakob, H. and Vizcaino, J.A. (2015). Making proteomics data accessible and reusable: current state of proteomics databases and repositories. *Proteomics,* **15**: 5-6, 930-949.

Petit, J.M., Ratinaud, M.H., Cordelli, E., Spano, M. and Julien, R. (1995). Mouse testis cell sorting according to DNA and mitochondrial changes during spermatogenesis. *Cytometry,* **19**: 4, 304-312.

Rafeeqi, T. and Kaul, G. (2013). Isolation and enrichment of type A spermatogonia from pre-pubertal buffalo (Bubalus bubalis) testis. *Andrologia,* **45**: 3, 195-203.

Reid, D.W., Shenolikar, S. and Nicchitta, C.V. (2015). Simple and inexpensive ribosome profiling analysis of mRNA translation. *Methods,* **91**: 69-74.

Robles, V. and Martinez-Pastor, F. (2013). Flow cytometric methods for sperm assessment. *Methods Mol Biol,* **927**: 175-186.

Rodriguez-Casuriaga, R., Folle, G.A., Santinaque, F., Lopez-Carro, B. and Geisinger, A. (2013). Simple and efficient technique for the preparation of testicular cell suspensions. *J Vis Exp,* 78,

Rodriguez-Casuriaga, R., Geisinger, A., Santinaque, F.F., Lopez-Carro, B. and Folle, G.A. (2011). High-purity flow sorting of early meiocytes based on DNA analysis of guinea pig spermatogenic cells. *Cytometry A,* **79**: 8, 625-634.

Rodriguez-Casuriaga, R., Santinaque, F.F., Folle, G.A., Souza, E., Lopez-Carro, B. and Geisinger, A. (2014). Rapid preparation of rodent testicular cell suspensions and spermatogenic stages purification by flow cytometry using a novel blue-laser-excitable vital dye. *MethodsX,* **1**: 239-243.

Rolland, A.D., Jegou, B. and Pineau, C. (2008). Testicular development and spermatogenesis: harvesting the postgenomics bounty. *Adv Exp Med Biol,* **636**: 16-41.

Romrell, L.J., Bellve, A.R. and Fawcett, D.W. (1976). Separation of mouse spermatogenic cells by sedimentation velocity. A morphological characterization. *Dev Biol,* **49**: 1, 119-131.

Rotgers, E., Cisneros-Montalvo, S., Jahnukainen, K., Sandholm, J., Toppari, J. and Nurmio, M. (2015). A detailed protocol for a rapid analysis of testicular cell populations using flow cytometry. *Andrology,* **3**: 5, 947-955.

Ryu, B.Y., Kubota, H., Avarbock, M.R. and Brinster, R.L. (2005). Conservation of spermatogonial stem cell self-renewal signaling between mouse and rat. *Proc Natl Acad Sci U S A,* **102**: 40, 14302-14307.

Schultz, N., Hamra, F.K. and Garbers, D.L. (2003). A multitude of genes expressed solely in meiotic or postmeiotic spermatogenic cells offers a myriad of contraceptive targets. *Proc Natl Acad Sci U S A,* **100**: 21, 12201-12206.

Shima, J.E., McLean, D.J., McCarrey, J.R. and Griswold, M.D. (2004). The murine testicular transcriptome: characterizing gene expression in the testis during the progression of spermatogenesis. *Biol Reprod,* **71**: 1, 319-330.

Shimizu, Y., Motohashi, N., Iseki, H., Kunita, S., Sugiyama, F. and Yagami, K. (2006). A novel subpopulation lacking Oct4 expression in the testicular side population. *Int J Mol Med,* **17**: 1, 21-28.

Shinohara, T., Orwig, K.E., Avarbock, M.R. and Brinster, R.L. (2000). Spermatogonial stem cell enrichment by multiparameter selection of mouse testis cells. *Proc Natl Acad Sci U S A,* **97**: 15, 8346-8351.

Simard, O., Gregoire, M.C., Arguin, M., Brazeau, M.A., Leduc, F., Marois, I., Richter, M.V. and Boissonneault, G. (2014). Instability of trinucleotidic repeats during chromatin remodeling in spermatids. *Hum Mutat,* **35**: 11, 1280-1284.

Simard, O., Leduc, F., Acteau, G., Arguin, M., Gregoire, M.C., Brazeau, M.A., Marois, I., Richter, M.V. and Boissonneault, G. (2015). Step-specific Sorting of Mouse Spermatids by Flow Cytometry. *J Vis Exp,* 106, e53379.

Song, Y., Liu, K.J. and Wang, T.H. (2014). Elimination of ligation dependent artifacts in T4 RNA ligase to achieve high efficiency and low bias microRNA capture. *PLoS One,* **9**: 4, e94619.

Soumillon, M., Necsulea, A., Weier, M., Brawand, D., Zhang, X., Gu, H., Barthes, P., Kokkinaki, M., Nef, S., Gnirke, A., Dym, M., de Massy, B., Mikkelsen, T.S. and Kaessmann, H. (2013). Cellular source and mechanisms of high transcriptome complexity in the mammalian testis. *Cell Rep,* **3**: 6, 2179-2190.

Spealman, P., Wang, H., May, G., Kingsford, C. and McManus, C.J. (2016). Exploring Ribosome Positioning on Translating Transcripts with Ribosome Profiling. *Methods Mol Biol,* **1358**: 71-97.

Vogel, C. and Marcotte, E.M. (2012). Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat Rev Genet,* **13**: 4, 227-232.

Wang, P.J., Page, D.C. and McCarrey, J.R. (2005). Differential expression of sex-linked and autosomal germ-cell-specific genes during spermatogenesis in the mouse. *Hum Mol Genet,* **14**: 19, 2911-2918.

Watson, J.V., Nakeff, A., Chambers, S.H. and Smith, P.J. (1985). Flow cytometric fluorescence emission spectrum analysis of Hoechst-33342-stained DNA in chicken thymocytes. *Cytometry,* **6**: 4, 310-315.

Wistuba, J., Stukenborg, J.-B. and Luetjens, M. (2007). Mammalian Spermatogenesis. *Functional Development and Embryology,* **1**: 99-117.

Xiao, Z., Zou, Q., Liu, Y. and Yang, X. (2016). Genome-wide assessment of differential translations with ribosome profiling data. *Nat Commun,* **7**: 11194.

Yadav, R.P. and Kotaja, N. (2014). Small RNAs in spermatogenesis. *Mol Cell Endocrinol,* **382**: 1, 498-508.

Yoshida, S., Sukeno, M., Nakagawa, T., Ohbo, K., Nagamatsu, G., Suda, T. and Nabeshima, Y. (2006). The first round of mouse spermatogenesis is a distinctive program that lacks the self-renewing spermatogonia stage. *Development,* **133**: 8, 1495-1505.

Zhao, H., Traganos, F., Dobrucki, J., Wlodkowic, D. and Darzynkiewicz, Z. (2009). Induction of DNA damage response by the supravital probes of nucleic acids. *Cytometry A,* **75**: 6, 510-519.

Zhou, K.W., Zheng, X.M., Yang, Z.W., Zhang, L. and Chen, H.D. (2009). Overexpression of CIRP may reduce testicular damage induced by cryptorchidism. *Clin Invest Med,* **32**: 2, E103-111.

Zhou, Q., Wang, M., Yuan, Y., Wang, X., Fu, R., Wan, H., Xie, M., Liu, M., Guo, X., Zheng, Y., Feng, G., Shi, Q., Zhao, X.Y., Sha, J. and Zhou, Q. (2016). Complete Meiosis from Embryonic Stem Cell-Derived Germ Cells In Vitro. *Cell Stem Cell,* **18**: 3, 330-340.

Zhuang, F., Fuchs, R.T., Sun, Z., Zheng, Y. and Robb, G.B. (2012). Structural bias in T4 RNA ligase-mediated 3'-adapter ligation. *Nucleic Acids Res,* **40**: 7, e54.

Zuccotti, P. and Modelska, A. (2016). Studying the Translatome with Polysome Profiling. *Methods Mol Biol,* **1358**: 59-69.

Zupanic, A., Meplan, C., Grellscheid, S.N., Mathers, J.C., Kirkwood, T.B., Hesketh, J.E. and Shanley, D.P. (2014). Detecting translational regulation by change point analysis of ribosome profiling data sets. *RNA,* **20**: 10, 1507-1518.

Concluding Remarks

This work combines two different strategies to explore the regulatory networks of spermatogenesis and the impact of a deregulation of this system in the context of male infertility. A comprehensive genetic characterization of two loci associated with phenotypes of SFF led to the identification of new candidate variants, showed the high conservation of key regulators of spermatogenesis throughout mammalian evolution and allowed the evaluation of the overall contribution of genetic diversity at these loci to diseases of the male reproductive system. This candidate-gene approach emphasized the complexity of the genetic architecture underlying the development SFF, suggesting that relevant mildly deleterious rare variants at spermatogenesis-related genes represent risk factors that collectively contribute to the phenotype. Additionally, with the methods developed here, functional genomics studies can be performed that will surely bring new insights on the functional impact of genetic variation at regulatory regions and protein coding sequence associated with defects of sperm cell development. Comparative studies on the regulatory processes governing spermatogenesis progression are a powerful tool to attribute functional relevance to the molecules involved in these pathways. Such work can be performed taking advantage of the Ho-FACS technique optimized in this work, as demonstrated by its successful application to the investigation of the spermatogenesis' translatome. Importantly, the work described here emphasizes the inherent challenge of tackling the etiology of complex diseases. Knowledge from different research fields must be integrated to allow a better understanding of the different layers regulating spermatogenesis, elucidate the underlying biology of this developmental process and how alterations can disturb this system and lead to disease.

An integrative view of spermatogenesis and its significance in the context of disease

An integrative approach that considers spermatogenesis as an assemblage of many micro-systems within a larger system is undoubtedly an extremely challenging task. Inter- and intracellular interactions as well as the cellular response to environmental stimuli comprise an enormous molecular network difficult to be taken into account as a whole. As systems biology can be essentially defined as the use of experimental and computational methods to study interactions between components of a biological system (Conesa and Mortazavi 2014), many parts of functional genomics fall within the purview of this field. There has been much controversy around the definition and integration of systems biology with functional genomics, addressed by many (Edwards and Thiele 2013; Schneider 2013; Conesa and Mortazavi 2014)

in interesting thorough discussions. Regardless, a synergistic cross-disciplinary approach seems to be mutually beneficial in the collection and functional inference of biological data. Assessing the biological relevance of perturbations to regulatory networks under different environmental conditions, and in the context of disease, can help to guide biomedical translational research in streamlining drug efficacy and safety testing (Valerio 2013; Sturla et al. 2014), biomarker discovery (Guingab-Cagmat et al. 2013) and designing and construction of novel therapeutics (Mol et al. 2015). This perspective is also shared amongst researchers in the field of male reproductive biology (Calvel et al. 2010; Com et al. 2014). Omics approaches for studying spermatogenesis have boosted in the past years with the realization that functional genomics holds the key to answer many of the questions so far difficult to address. As elegantly stated by Com and co-workers (Com et al. 2014): “If we are to understand pathological disorders of the human testis and their origins we will need to unravel the complete set of transcriptional regulation mechanisms leading to the coordinated expression of a myriad of functional products and identify the protein network interactions occurring during spermatogenesis”. In this view, the strategy adopted in this work describes the effort to address different regulatory layers of spermatogenesis towards a better understanding of disease-causing alterations. First, all efforts to assess the distribution of potentially deleterious variants in genes relevant for spermatogenesis in different populations will contribute to a better understanding of the epidemiology of SFF. Second, the application of the tools described here to other germ cell types that may be isolated by Ho-FACS in different mammalian species has the potential to unravel the intricate mechanisms of translational regulation occurring during gamete formation. Moreover, the protocol for Ho-FACS isolation of mammalian germ cells can be used to collect starting material for high-throughput omics studies focusing on other layers of spermatogenesis regulation. An integrative omics strategy using these datasets and others from studies of gene and protein expression in spermatogenesis already available in online repositories is the next step to bring new insights into the mechanisms underlying germ cell development. This knowledge is crucial for the identification of new genes that may be associated with SFF and to provide functional annotation to genome-wide association studies, and ultimately may contribute to the improvement of the diagnosis and treatment of male infertility.

Currently, the main solution to surpass male infertility due to azoospermia relies on the use of patient gametes with incomplete development, when available, for intracytoplasmic sperm injection (ICSI). In fact, assisted reproduction has already been performed using round spermatids (Tanaka et al. 2015) or abnormal sperm cells, such as those collected from patients

with easily decapitated sperm syndrome (Emery et al. 2004). It is however noteworthy the significance of the paternal contribution with DNA, RNA, patterns of imprinting for proper embryogenesis has been already demonstrated (Ostermeier et al. 2004; Gur and Breitbart 2006; Rassoulzadegan et al. 2006) and the fact that assisted reproductive techniques (ARTs) may be associated with higher incidence of some genetic diseases, such as Angelman syndrome. Also, it was recently shown that histone retention alters gene expression in the embryo (Ihara et al. 2014), suggesting that histone to protamine transition during spermiogenesis affects normal embryonic development. Interestingly, the aforementioned studies have shown successful pregnancies giving birth to apparently normal offspring without signs of any developmental defects, suggesting that at least to some extent, defects in spermiogenesis have little effect on embryogenesis. Nonetheless, caution must be taken when facilitating the transmission of genetic material from low quality gametes since they may carry mildly deleterious variants that would have otherwise been eliminated from the population by natural selection. While the psychological and social burden of infertility is not one to be ignored, a better understanding of the full long term consequences of ARTs both at the individual and the species level is urgently needed. Efforts must be put into the development of diagnostic tools that allow for patient-specific identification of the cause of disease and allow an informed and adequate treatment response. In this respect, integrative genomics strategies have already proven a valuable approach with significant clinical applications. Researchers are developing a multiplex assay called Fertichip™ to detect the presence/absence of post-meiotic germ cell biomarkers identified from seminal fluid from healthy donors (Rolland et al. 2013). This assay will facilitate prediction of spermatozoa retrieval by surgery and therefore represents a major advance for counseling and managing NOA patients (Com et al. 2014). Similarly, data generated by the tools described in chapter 2 holds the promise to unveil many new contributors to the male infertility phenotype and help to guide clinicians towards diagnostics and solutions in a patient-specific manner.

Literature Cited

Calvel P, Rolland AD, Jegou B, Pineau C. 2010. Testicular postgenomics: targeting the regulation of spermatogenesis. *Philosophical transactions of the Royal Society of London Series B, Biological sciences* **365**: 1481-1500.

Com E, Melaine N, Chalmel F, Pineau C. 2014. Proteomics and integrative genomics for unraveling the mysteries of spermatogenesis: the strategies of a team. *Journal of proteomics* **107**: 128-143.

Conesa A, Mortazavi A. 2014. The common ground of genomics and systems biology. *BMC systems biology* **8 Suppl 2**: S1.

Edwards LM, Thiele I. 2013. Applying systems biology methods to the study of human physiology in extreme environments. *Extreme physiology & medicine* **2**: 8.

Emery BR, Thorp C, Malo JW, Carrell DT. 2004. Pregnancy from intracytoplasmic sperm injection of a sperm head and detached tail. *Fertility and sterility* **81**: 686-688.

Guingab-Cagmat JD, Cagmat EB, Hayes RL, Anagli J. 2013. Integration of proteomics, bioinformatics, and systems biology in traumatic brain injury biomarker discovery. *Frontiers in neurology* **4**: 61.

Gur Y, Breitbart H. 2006. Mammalian sperm translate nuclear-encoded proteins by mitochondrial-type ribosomes. *Genes Dev* **20**: 411-416.

Ihara M, Meyer-Ficca ML, Leu NA, Rao S, Li F, Gregory BD, Zalenskaya IA, Schultz RM, Meyer RG. 2014. Paternal poly (ADP-ribose) metabolism modulates retention of inheritable sperm histones and early embryonic gene expression. *Plos Genet* **10**: e1004317.

Mol M, Raj Bejugam P, Singh S. 2015. Synthetic biology at the interface of functional genomics. *Briefings in functional genomics* **14**: 180-188.

Ostermeier GC, Miller D, Huntriss JD, Diamond MP, Krawetz SA. 2004. Reproductive biology: delivering spermatozoan RNA to the oocyte. *Nature* **429**: 154.

Rassoulzadegan M, Grandjean V, Gounon P, Vincent S, Gillot I, Cuzin F. 2006. RNA-mediated non-mendelian inheritance of an epigenetic change in the mouse. *Nature* **441**: 469-474.

Rolland AD, Lavigne R, Dauly C, Calvel P, Kervarrec C, Freour T, Evrard B, Rioux-Leclercq N, Auger J, Pineau C. 2013. Identification of genital tract markers in the human seminal plasma using an integrative genomics approach. *Hum Reprod* **28**: 199-209.

Schneider MV. 2013. Defining systems biology: a brief overview of the term and field. *Methods in molecular biology* **1021**: 1-11.

Sturla SJ, Boobis AR, FitzGerald RE, Hoeng J, Kavlock RJ, Schirmer K, Whelan M, Wilks MF, Peitsch MC. 2014. Systems toxicology: from basic research to risk assessment. *Chemical research in toxicology* **27**: 314-329.

Tanaka A, Nagayoshi M, Takemoto Y, Tanaka I, Kusunoki H, Watanabe S, Kuroda K, Takeda S, Ito M, Yanagimachi R. 2015. Fourteen babies born after round spermatid injection into human oocytes. *Proc Natl Acad Sci U S A* **112**: 14629-14634.

Valerio LG, Jr. 2013. Predictive computational toxicology to support drug safety assessment. *Methods in molecular biology* **930**: 341-354.

Appendix

SUPPLEMENTARY DATA

Paper II - Rare double sex and mab-3-related transcription factor 1 regulatory variants in severe spermatogenic failure

Extended Materials and Methods

Samples

Samples of male individuals with idiopathic NOA were collected for routine molecular diagnosis of male infertility, during a period of over 10 years. Most DNA samples were extracted by salting out from peripheral blood leukocytes and samples were stored at -20°C. Control samples were collected over the last 4 years at IPATIMUP and at a fertility clinic. In the first case, genomic DNA was extracted from FTA cards using the GENERATION® Capture Card Kit (Gentra Systems, Minneapolis, USA) according to the manufacture's guidelines for Purifying DNA from difficult-to-elute dried blood spots. For the remaining samples DNA was extracted from blood using Citogene® DNA Blood Kit (Citomed, Lisbon, Portugal). Our previous analysis of genome-wide SNPs did not detect significant homozygosity-by-descent in the Portuguese infertile patients (Lopes *et al.* 2013) and thus there is no evidence of high levels of inbreeding in this cohort. Cases and controls were geographically matched.

The ancestry of a NOA patient harboring the c.354+38_insG and juxtaposed poly-T as well as the rs144122237 5'UTR variant was determined by multiplex of 46 ancestry informative markers (AIMs). This panel comprises a set of insertion-deletion polymorphisms selected to efficiently infer biogeographical ancestry and admixture proportions of four different origins: African, European, East Asian and Native American (Pereira *et al.* 2012). Biogeographical ancestry was assigned using Snipper 2.0 (http://mathgene.usc.es/snipper/) by calculating the profile's likelihood of belonging to each of the populations and performing a principal component analysis considering HGDP-CEPH samples as reference (Supplementary Fig. 1).

DMRT1 sequence analysis

The *DMRT1* genomic reference sequence (NM_021951.2) was obtained with annotated exons, introns, polymorphisms (ENSG00000137090; Ensembl database) and repeat elements (chr9:841,690-969,090; GRCh37/hg19 assembly; UCSC Genome Browser). From the 155 samples collected from patients, full gene sequences were obtained for 132 NOA but due to poor DNA quality for 23 only partial sequences were obtained. Altogether, we sequenced 139 patients for exons 1, 140 for exon 2, 138 for exon 3, 145 for exon 4 and 153 for exon 5. Nonetheless, variants' frequencies were calculated independently using all available sequences for each amplified fragment.

Rare variants at ≤1% in European populations and overrepresented in patients (see details bellow) were genotyped in Portuguese fertile and normozoospermic controls.

The presence of patient specific size variants in introns 1 (c.354+38_insG and juxtaposed poly-T) and 3 (rs59834456) was tested in 354 (53 normozoospermic 301 fertile males) and 102 (62 fertile and 40 normozoospermic) Portuguese controls, respectively, by standard PCR fragment amplification followed by agarose or polyacrylamide gel electrophoresis.

The promoter variant (c.-223_-219CGAAA>T), 5'-UTR variant (rs144122237) and synonymous substitution in exon 1 (rs3739583) were genotyped by PCR amplification and Sanger sequencing of a fragment containing both variants, following the same procedure as for patients, and using the primers described in supplementary table 1. This approach allowed us to calculate the allele frequencies for the other variants present in this region (rs3739584 and rs3739583).

A SNaPshot Multiplex reaction (Supplementary Fig.2) was designed in order to, in a single reaction, genotype Portuguese controls for two novel variants (c.355-6T>C and c.823-64_823-62delATT) and 4 rare variants overrepresented in patients when compared to European populations (rs55905583; rs146975077; rs34946058; rs200423545). Supplementary Table 2 summarizes sequences of primers and probes used, fragment sizes and allele reading obtained by the aforementioned SNapShot reaction. DNA fragments were amplified by multiplex PCR, SNaPshot was performed using the ABI PRISM® SNaPshot™ Multiplex Kit (Applied Biosystems®, Life Technologies, Carlsbad, California, USA) and products were run on an ABI PRISM 3130xl Genetic Analyzer (Applied Biosystems®, Life Technologies, Carlsbad, California, USA). Peak calling was obtained and analyzed using the GeneMaper Software (Applied Biosystems®, Life Technologies, Carlsbad, California, USA).

In silico and statistical analysis

We initially set a control group of populations of European ancestry by calculating the total number of observations of each allele for all variants found in: a) 1000 Genomes European (EUR) and European descent (CEU) populations; b) ESP European-American (EUR-AM) population; and c) CLINSEQ European descent (CSAgilent) population. While the 1000 Genomes Project aimed at detecting genetic variants with frequencies of at least 1% in unphenotyped individuals from several different populations worldwide, both the ESP initiative and the CLINSEQ project comprise a catalogue variants found in individuals well characterized for heart, lung and blood disorders or heart disease, breast cancer, and hearing loss respectively. Allele frequency of variants found in patients was compared to this established

European population in order to select rare variants to be genotyped in the Portuguese control group of fertile and normozoospermic

In silico assessment of potential regulatory roles of variants requires a cautious analysis of the bioinformatics predictions. Interpretation of the scores obtained using the Human Splicing Finder (HSF) v.2.4.1 software was performed following the guidelines described in (Desmet *et al.* 2009). The software provides a consensus value (CV) for reference and mutant splice sites that is defined based on over 400.000 naturally occurring splice sites. In the vast majority of the genes, active splice sites have CVs over 70 with a mean value for 5'ss of 87.53 ± 8.34 and for 3'ss of 86.81±6.33 (Desmet *et al.* 2009). HSF also predicts binding of exonic splicing enhancers based on the ESE Finder software (Cartegni *et al.* 2003) where scores above the experimentally calculated threshold for each enhancer are considered to be significant.

References

Cartegni L, Wang J, Zhu Z, Zhang MQ, Krainer AR. (2003) ESEfinder: A web resource to identify exonic splicing enhancers. *Nucleic acids research* 31, 3568-71.

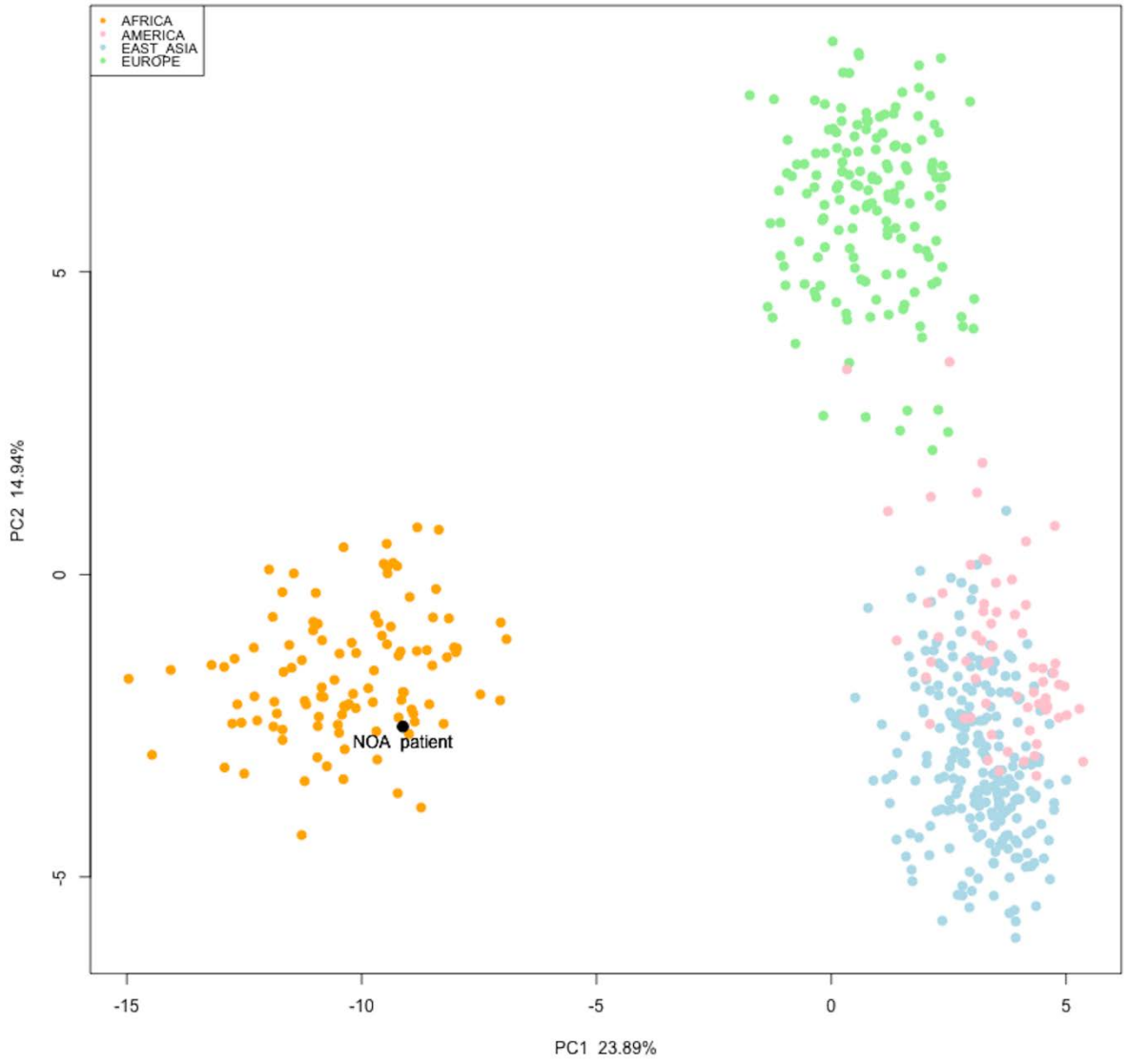
Desmet FO, Hamroun D, Lalande M, Collod-Beroud G, Claustres M, Beroud C. (2009) Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic acids research* 37, e67.

Lopes AM, Aston KI, Thompson E, Carvalho F, Goncalves J, Huang N *et al.* (2013) Human Spermatogenic Failure Purges Deleterious Mutation Load from the Autosomes and Both Sex Chromosomes, including the Gene DMRT1. *Plos Genet* 9.

Pereira R, Phillips C, Pinto N, Santos C, dos Santos SE, Amorim A, Carracedo A, Gusmao L. (2012) Straightforward inference of ancestry and admixture proportions through ancestry-informative insertion deletion multiplexing. *Plos One* 7, e29684.

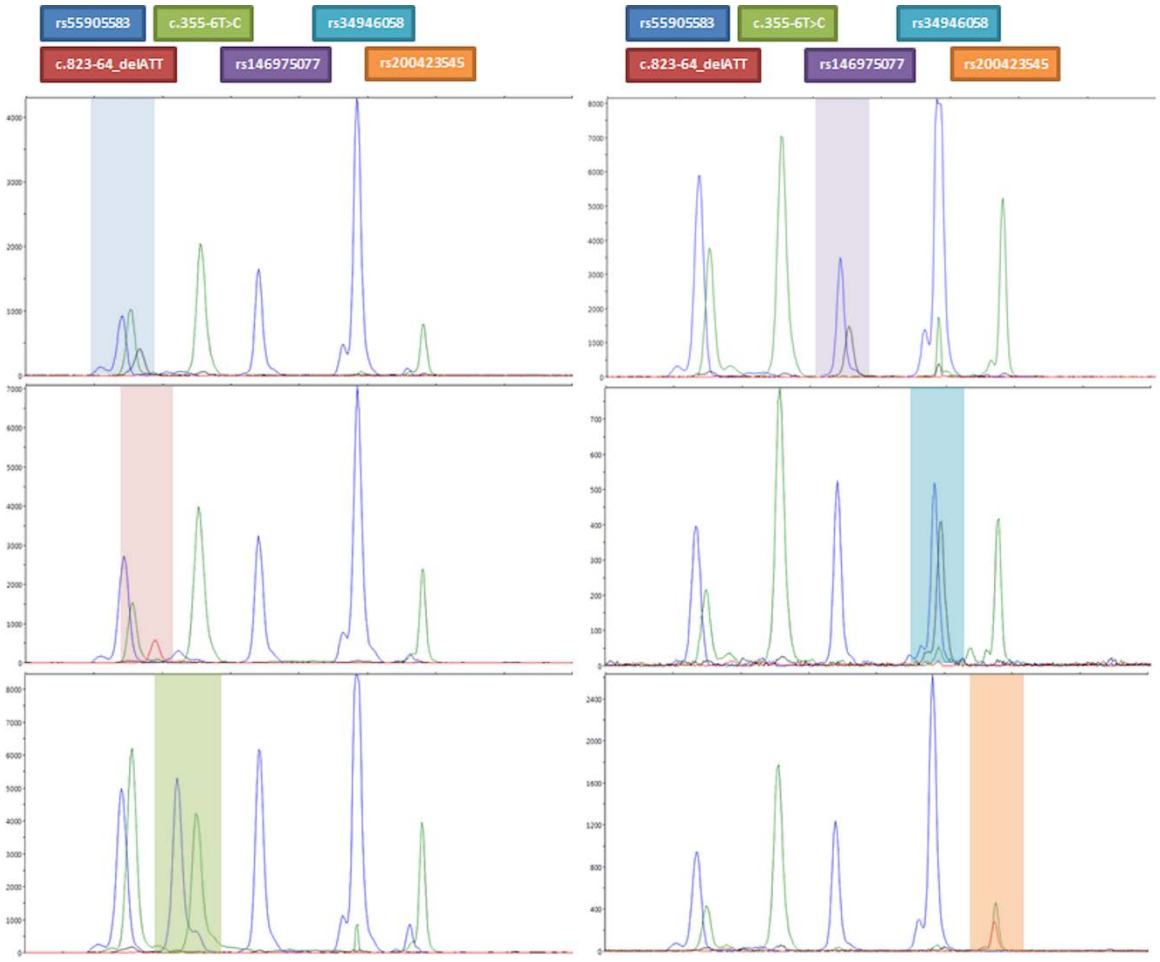
Supplementary Figure 1. Principal component analysis of reference populations and NOA patient tested using AIMs.

The rs144122237 5'UTR variant found in one NOA patient is present at 1.8% in African populations. In order to understand if its detection in our cohort was due to admixture with African populations, we genotyped this individual for a panel of 46 AIMs. This plot shows the patient's profile clustering together with African samples, indicating an African ancestry, obtained using Snipper 2.0 (http://mathgene.usc.es/snipper/).



Supplementary Figure 2. SNaPshot profiles of the heterozygous state for all variants included in the multiplex reaction.

SNaPshot was performed in those patients heterozygous for the variants tested in order to determine the profile for the alternative allele of each locus. The SNaPshot profiles in this figure represent the heterozygous state for one variant marked by the bin (shadowed box) overlapping the peaks. Filled colored boxes at the top of the figure identify the variants genotyped in this multiplex with colors corresponding to the assigned bins. Profiles were generated using the Peak Scanner Software (Applied Biosystems®, Life Technologies, Carlsbad, California, USA)



Supplementary Figure 3. Snapshot of the alignment of mammalian and avian DMRT1 sequences

Aligning the protein sequences of several mammalian species and the chicken, a region immediately downstream of the DMRT1-like domain (purple filled box), strikes out as highly conserved and is almost free of missense variants (green box). This alignment was obtaining applying the ClustalW algorithm available with the Geneious v5.5.8 software. Amino acids are colored when they match at least 75% of the sequences. Red filled boxes represent missense variants annotated in the Ensembl database predicted by at least one software (SIFT or PolyPhen) as deleterious.

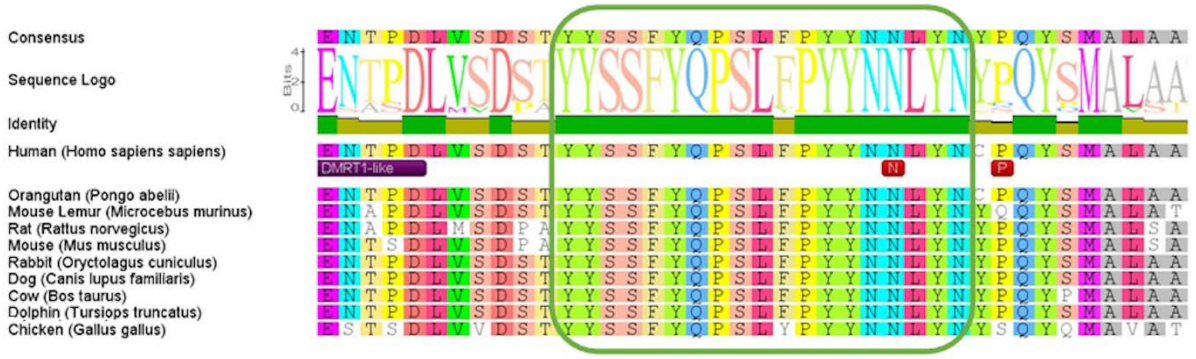


Table S1. Primers designed for *DMRT1* amplification and sequencing

Region	Primer type	3'- 5' Seq	Size (bp)	Tm (°C)	Product Size (bp)	Application
Promoter	Fwd_II*	GGTGGTTTGCACCTCCTCTA	20	60.11	328	Allele specific amplification of promoter variants
	E1_insertion.spec_REV	CGCGCGACTCAGAACTTT	18	60.29		
Promoter	Fwd_II*	GGTGGTTTGCACCTCCTCTA	20	60.11	324	Allele specific amplification of promoter variants
	E1_deletion.spec_REV	CGCGCGACTCAGAACAG	17	60.45		
Promoter	Fwd_II*	GGTGGTTTGCACCTCCTCTA	20	60.11	554	Amplification of promoter region and 5'-UTR in controls
	DMRT1_E1Promoter_REV	CTGAATGCCTCGTCGTTG	18	58.89		
Exon 1	Fwd*	CCGGGAATGTTCTGAAAAGTAT	22	59.37	972	Amplification of Exon 1 in patients
	Rev	GGGGGACTTCATTCAAGAAAC	21	59.80		
Exon 1	Fwd_II*	GGTGGTTTGCACCTCCTCTA	20	60.11	1257	Amplification of Exon 1 in patient with promoter variants
	Rev	GGGGGACTTCATTCAAGAAAC	21	59.80		
Intron 1	NESTED_DMRT1_E1_FWD	CAACCTGATCGCCGAGAG	18	60.51	152	Amplification of the region containing the intron 1 poly-T in controls
	NESTED_DMRT1_E1_REV	CAAGATCGCGCCACTACAC	19	60.43		
Exon 2	Fwd*	GTTTCTCAGCTTTGCACATCA	21	58.13	692	Amplification of Exon 2 in patients
	Rev	CAAGGATATTTAGTTCCCACAAGG	24	60.12		
Exon 3	Fwd*	CCTTGCTCCGCAGGTCTT	18	61.51	634	Amplification of Exon 3 in patients
	Rev	CAATTCCTTGCAGCCAACC	19	61.57		
Intron 3	DMRT1_I3_FWD	GGAAACCAGTGGCAGGTATG	20	60.38	266	Amplification of the region containing the intron 3 size variant rs59834456 in controls
	DMRT1_I3_REV	ATTCCTTGCAGCCAACCTTT	20	60.99		
Exon 4	Fwd	GCACCAAATGAACAGAAGGA	20	58.70	764	Amplification of Exon 4 in patients
	Rev*	TTAGACACAGCTAATGACCCAATAC	25	58.71		
Exon 5	Fwd*	GAGCGTCACTTTCTTTGTTGT	21	57.13	1262	Amplification of Exon 5 in patients
	Rev	TTTCCTGTTTAATACCGCTCAC	22	58.32		
Exon 5	Fwd	GAGCGTCACTTTCTTTGTTGT	21	57.13	582	Amplification of Exon 5 in patients with deletions upstream the CDS of exon 5
	Rev_II*	AGAGGCACACAAATGGCTTC	20	60.26		
Exon 5	Fwd_Seq*	CTGGCTTGGTTTCCCTCTC	19	59.79	.	Internal primer for sequencing exon 5
Exon 5	Rev_Seq*	ACTGCTCACTCGTCCTCCTC	20	59.58	.	Internal primers for sequencing exon 5

*Primer used for sequencing

Table S2. Summary of novel and rare variants tested in Portuguese patients and controls

Variant	Patients					Controls				
	Code	Histology	Testicular volume (R/L)	Hormone levels	Allele Freq.	Normozoospermic	Fertile	Total	Method*	Allele Freq.**
c.-223_-219CGAAA>T	Y2446	MA	N/N	LH:NA; FSH:16.8; T:3.32	1/278	56	301	357	Sanger Sequencing	0/714
c.354+38_insG(T)$_{n>10}$	Y2386	.	.	.	1/276	56	301	357	PCR+PAGE	0/714
c.355-6T>C	Y0209	HP	N/N	LH:5.8; FSH:9.1; T:NA	1/280	73	277	350	SNaPshot Multiplex	0/700
c.823-64_delATT	Y1813	.	.	.	1/290	65	265	330	SNaPshot Multiplex	0/660
rs59834456	Y1264	SCOS	.	.	5/274	40	62	102	PCR+PAGE	10/102***
	Y1530	.	.	.						
	Y1961	SCOS	.	.						
	Y2094	.	.	.						
	Y2758	SCOS	.	.						
rs144122237	Y2386	.	.	.	1/276	56	291	347	Sanger Sequencing	0/694
rs146975077	Y2098	SCOS	.	.	1/304	72	277	349	SNaPshot Multiplex	1/698
rs200423545	Y1268	MA	H/H	LH:7.9; FSH:11.5; T:3.0	1/276	73	272	345	SNaPshot Multiplex	4/690
	Y1456	HP	.	.						
rs34946058	Y1816	SCOS	.	.	3/276	72	276	348	SNaPshot Multiplex	3/696
	Y1880	.	.	LH:4.58; FSH:12.38; T:4.39						
	Y1989	.	.	.						
rs376518776	Y3155	MA	.	.	1/278	52	217	269	Sanger Sequencing	0/534
rs55905583	Y1414	SCOS	H/N	.	1/280	73	277	350	SNaPshot Multiplex	4/700
	Y1915	HP	N/N	LH:2.13; FSH:4.73; T:4.66						

MA: Maturation Arrest
HP: Hypoplasia
SCOS: Sertoly cell-only syndrome
N: Normal
H: Hypoplastic
LH: Luteinizing hormone
FSH: Follicle stimulating hormone
T: Testosterone
NA: Not available
PAGE: Polyacrylamide gel electrophoresis
* all variants were confirmed by Sanger sequencing
** only rs59834456 was detected in normozoospermic controls
*** frequency of carriers

Table S3. SNaPshot multiplex design

Fragment	PCR primer Name	PCR primer (5'-3')	Tm (°C)	Product Size (bp)	Variant	SBE primer Name	SBE (5'-3')	Final Size (bp)	Ref Seq	Reading
2	SNaP_F2_Fwd = DMRT1 E2_Fwd*	GTTTCTCAGCTTTGCACATCA	59.06	223	rs55905583	SBE_rs55905583_FWD	AGGATGACTCATTGTCGTGT	20	G/C	G/C
	SNaP_F2_Rev	GCTGATACCCAATTCCTCCTC	59.92		c.355-6T>C	SBE_c.355-6T>C_REV	aagtctgacaaCCTTCTCAGGGCCACCTGGA	31	T/C	A/G
3	SNaP_F3_Fwd*	TGTGGAGAACACACCTGACC	59.55	241	rs146975077	SBE_rs146975077_FWD	gtcgtgaaagtctgacaaCCTGTGAAGAACAGCCTTCG	38	G/C	G/C
	SNaP_F3_Rev	TCATACCTGCCACTGGTTTC	58.57		rs34946058	SBE_rs34946058_REV	gtgccacgtcgtgaaagtctgacaaTGACCAGGCACATAAGGTCC	45	C/G/T	G/C/A
4	SNaP_F4_Fwd*	AGGCTATTCCCTTAATTTGCTTT	58.82	425	c.823-64_delATT	SBE_c.823-64_delATT_REV	acaaCATGTCACAACTAGTACAAT	24	ATT/-	A/T
	SNaP_F4_Rev	TCTCCATGTTCTTCATCTGCTT	58.39							
5	SNaP_F5_Fwd	GCCACTTTTAACATTACTCCCTTT	59.07	220	rs200423545	SBE_rs200423545_REV	taggtgccacgtcgtgaaagtctgacaaAAGTGAGGTGAGAG	48	T/A	A/T
	SNaP_F5_Rev*	CTGCTCACTCGTCCTCCTC	58.63							

* Sequencing primer

SBE: Single Base Extension

Table S4. *DMRT1* variants retrieved from Ensembl database

Large table. See: http://onlinelibrary.wiley.com/doi/10.1111/andr.12063/suppinfo

Table S5. TFBSs predicted for reference or altered human sequences in the conserved motif in the *DMRT1* promoter sequence

Input Seq	Software	TFBSs	score
Ref Human	MatInspector	Heat shock factor 2	0.977[a]
		Heat shock factor 1	0.767[a]
		SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 3	0.993[a]
		Myelin transcription factor 1-like, neuronal C2HC zinc finger factor 1	0.986[a]
	TFSearch	HSF2	92.3[b]
		GATA-2	89.3[b]
		HSF1	86.2[b]
Human w/ Ancestral	MatInspector	Heat shock factor 2	0.977[a]
		SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 3	0.993[a]
		Doublesex and mab-3 related transcription factor 3	0.829[a]
		MyT1 zinc finger transcription factor involved in primary neurogenesis	0.901[a]
	TFSearch	None	.
Patient	MatInspector	Heat shock factor 2	0.977[a]
		SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 3	0.993[a]
		Ecotropic viral integration site 1 encoded factor, amino-terminal zinc finger domain	0.849[a]
		Progesterone receptor binding site, IR3 sites	0.89[a]
	TFSearch	None	.

[a] scale 0-1

[b] scale 0-100

Table S6. TFBS motifs overlaping with variants in the 5'UTR

dbSNP ID	Matrix	Detailed Matrix Information	Strand	Core similarity	Matrix similarity	Sequence
rs144122237	V$RFX3.02	Regulatory factor X, 3 (secondary DNA binding preference)	-	0.847	0.922	tctctcgcctgGAGActgctg
rs3739584	V$SMARCA3.02	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 3	+	1	0.987	tcgcACTTctc
	V$IR2_NGRE.01	Repressive binding sites for glucocorticoid receptor (IR2)	-	0.759	0.891	tgCCCCtaggagaag
	V$GCM1.03	Glial cells missing homolog 1 (secondary DNA binding preference)	-	1	0.837	tggtgCCCCtaggag

Table S7. *DMRT1* cDNA variants detected in large genome sequencing projects retrieved from Ensembl database.

Large table. See: http://onlinelibrary.wiley.com/doi/10.1111/andr.12063/suppinfo

SUPPLEMENTARY DATA

Paper III - The Mutational Spectrum of *WT1* in Male Infertility

Supplementary table 1 - WT1 primers used in this study

Primer	Sequence	Length (bp)	Tm (°C)	GC %	Product Length
Exon 1A_F kohler (Kohler, Pienkowski et al. 2004)	CAGCGCTGAACGTCTCCA	18	66,72	61,11	573 bp
Exon 1A_R kohler (Kohler, Pienkowski et al. 2004)	GGGTGTCCTAGAGCGGAGAG	20	65,14	65,00	
WT1_Exons 2+3_F	AGCCCCAGACAGATAACA	18	58,08	50,00	1505 bp
WT1_Exons 2+3_R	TTCCTCCCAGTAAAGACC	18	56,47	50,00	
WT1_Exons 4+5_F	CTGGAAAATGTGGAGGCT	18	60,21	50,00	1386 bp
WT1_Exons 4+5_R	TGCTACCCTGATTACCCA	18	59,00	50,00	
WT1_Exon 6_F	GCCTCATCTCATCTGGAAGT	20	60,81	50,00	1020 bp
WT1_Exon 6_R	GGTGTCCCTGATGTTAAAGG	20	60,78	50,00	
WT1_Exon 7_F	CCTCAAGACCTACGTGAATGT	21	60,59	47,62	458 bp
WT1_Exon 7_R	ACTTTCTCTCTACCACTCTGCTC	23	59,72	47,83	
WT1_Exon 8_F	CTAACAAGCTCCAGCGAAGT	20	61,13	50,00	697 bp
WT1_Exon 8_R	TCATGCCTCACCCTTAGATT	20	61,03	45,00	
WT1_Exon 9_F	TAGCAGTGGGCTGATGATAC	20	60,27	50,00	734 bp
WT1_Exon 9_R	GTAGGGACCTGGCTTATCTCT	21	60,12	52,38	
WT1_Exon 10_F	GTTAGCTCAGGGACAGAATGA	21	60,82	47,62	772 bp
WT1_Exon 10_R	TGACCTCGGGAATGTTAGAC	20	61,49	50,00	

Primers for amplification and sequencing were designed in Primer3 v.0.4.0 (http://bioinfo.ut.ee/primer3-0.4.0/) using the latest version of the human genome assembly (GRCh37).

Supplementary table 2 – WT1 missense variants found in large scale sequencing projects (Exome Sequencing Project – ESP; 1000Genomes – 1KG and HapMap) with global minor allele frequencies (estimated using allele frequencies across European Americans and African Americans – ESP or across all 1000G Phase I populations).

Residue	Exon	Variation ID	Source	MAF (Global)	Alleles	Residues	Polyphen
193	E1	rs377072761	ESP	0.0008	C/**A**	A, S	0.034
199	E1	rs9332973	HapMap - CEU	0.0130	C/**T**	A, T	0.496
215	E1	rs373935628	ESP	0.0008	G/**T**	Q, K	0.479
249	E2	rs2234584	ESP	0.0006	G/**A**	P, S	0.034
250	E2	rs142653301	ESP	0.0008	T/**C**	M, V	0.020
272	E3	rs138073760	ESP	0.0002	G/**A**	T, I	0.133
282	E3	rs368452754	ESP	0.0008	C/**A**	A, S	0.277
336	E6	rs371021920	ESP	0.0008	C/**T**	S, N	0.057
350	E6	rs142059681	ESP	0.0008	A/**G**	C, R	0.987
362	E6	rs150194429	ESP	0.0008	A/**C**	F, C	0.994
380	E7	rs147241955	1KG	0.0010	C/**T**	R, Q	0.945
381	E7	rs142937387	1KG	<0.001	G/**C**	S, W	0.998
413	E7	rs373176048	ESP	0.0008	C/**A**	R, M	0.999
430	E8	rs144788858	ESP	0.0008	C/**T**	R, Q	0.096
485	E10	rs139893274		<0.01	C/**T**	R, Q	0.684
			1KG				
504	E8	rs369940913	ESP	0.0008	G/**C**	N, K	0.925

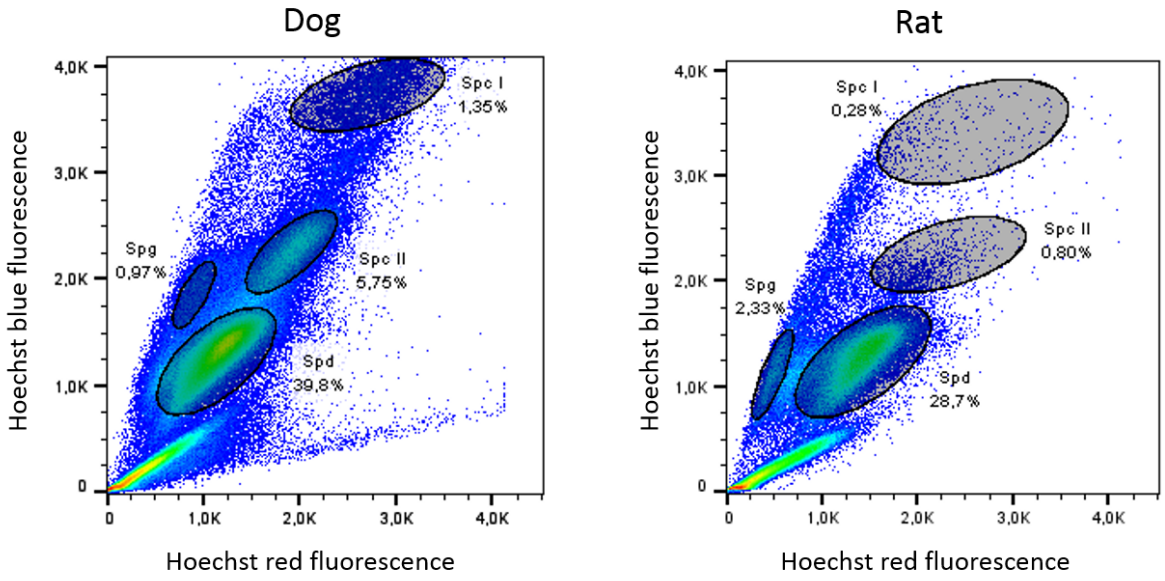
SUPPLEMENTARY DATA

Paper IV - Multispecies purification of testicular germ cells

(in revision)

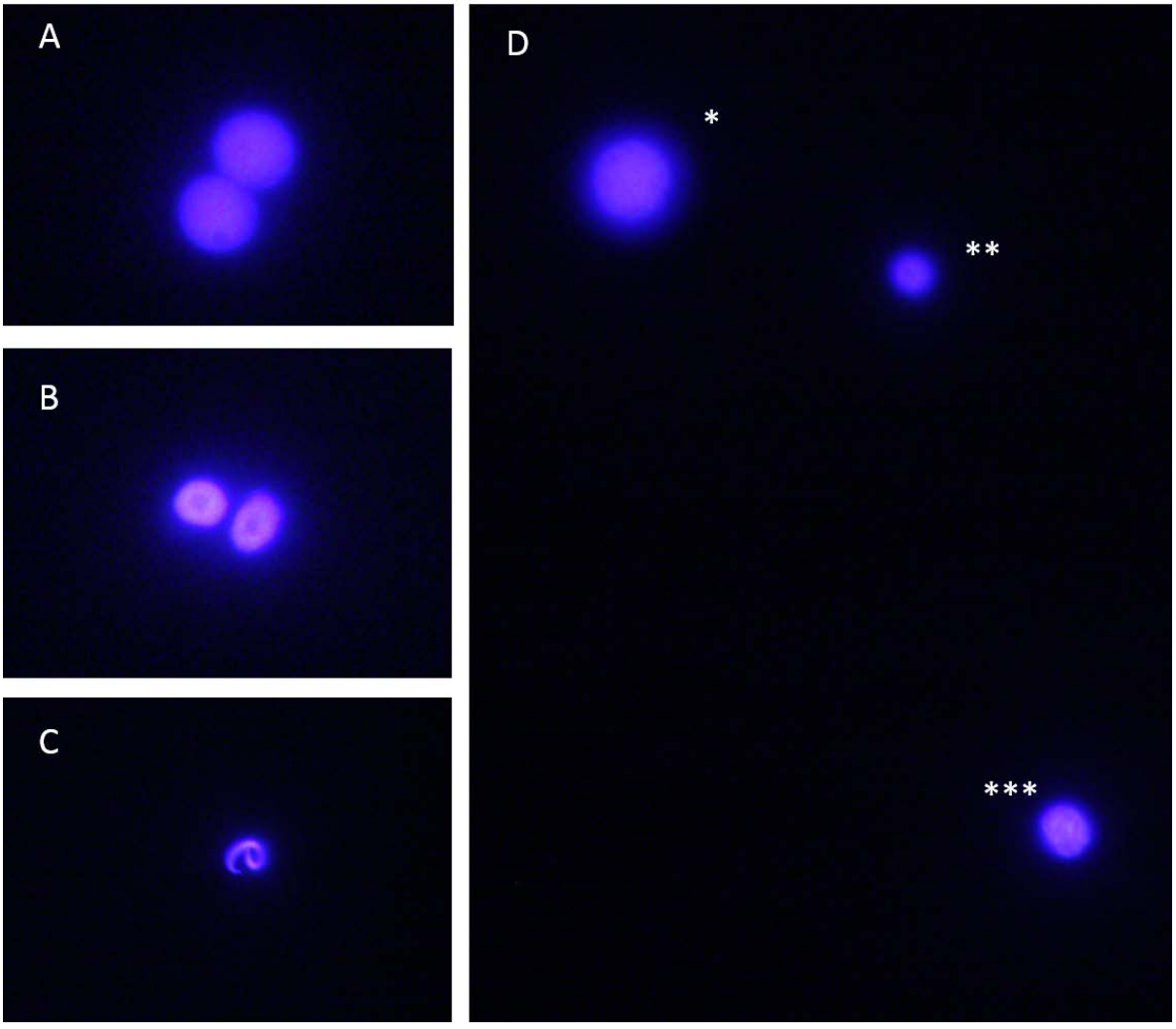
Supplementary Figure 1. Ho-FACS plots of cell suspensions obtained using a generalized testis dissociation protocol.

Sample preparation has direct implications in the success and results obtained by flow cytometry. In this figure, plots reflect measurements of Hoechst fluorescence of stained germ cells isolated from testes of dog and rat, using a generalized dissociation multispecies protocol, optimized for mouse. For both samples we observe only two of the four expected populations, indicating that some germ cell-types were absent from the initial single-cell suspension. Round circles represent the gates defined based on the expected location of cell clusters in respect to Ho fluorescence. Percentages indicate the proportion of cells within the gates in relation to the total number of live cells. Spg: spermatogonia; Spc I: primary spermatocytes; Spc II: secondary spermatocytes; Spd: spermatids. Plots were obtained using FlowJo® software v10 (Tree Star Inc.).



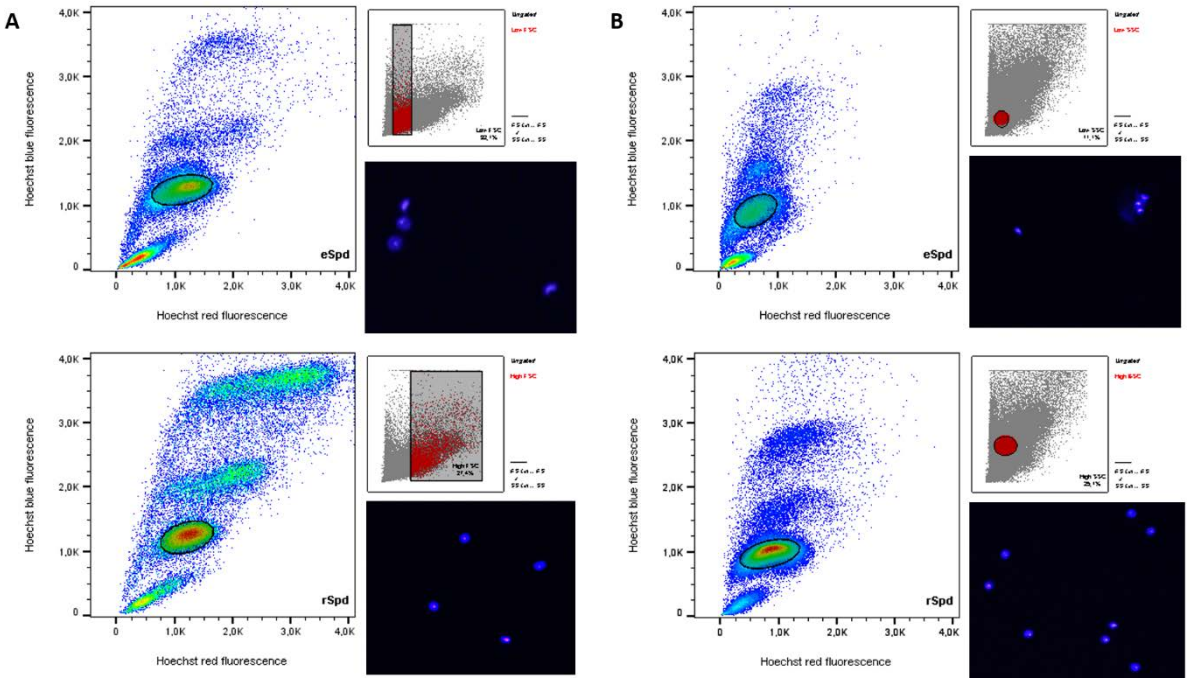
Supplementary Figure 2. Microscopic identification of frog germ cells sorted in gate "Spg".

Based on chromatin structure (blue fluorescence) and cell shape, we identified frog germ cells in different developmental stages sorted in the gate "Spg". This population comprised a mixture of secondary spermatocytes (A), round spermatids (B; D: **), spermatozoa (C), spermatogonia (D:***) and primary spermatocytes (D:*), indicating that Ho-FACS was inefficient in the isolation of specific germ cell-types in the allotetraploid frog. Hoechst fluorescence was visualized after FACS in a light microscope using a UV lamp under a 63X magnification lens.



Supplementary Figure 3. Optimization of a gating strategy to isolate round and elongating spermatids.

In order to discriminate between round (rSpd) and elongating spermatids (eSpd) we defined the parent gates (circles and squares with cells labeled red) to reflect differences in cell shape (A) or complexity (B). Gates for sorting were then defined by the expected pattern of Hoechst blue/red fluorescence for spermatids. Cell populations gated for high or low FSC were enriched 62% for rSpd and 84% for eSpd respectively (A). Within a range of low FSC, gating for higher or lower SSC increased the enrichment to 86% and 92% of rSpd and eSpd, respectively, in the sorted population (B). Morphology of sorted cells was evaluated based on Hoechst fluorescence and images were acquired by light microscopy with a UV lamp (16X magnification lens.). Plots were generated using FlowJo® software v10 (Tree Star Inc.)



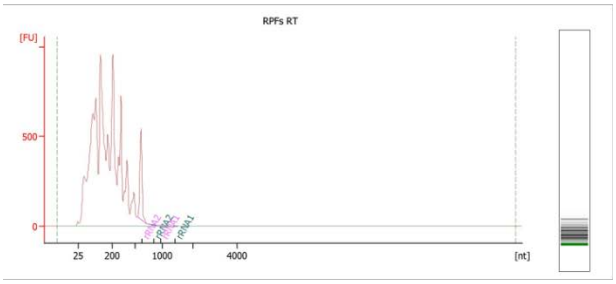
SUPPLEMENTARY DATA

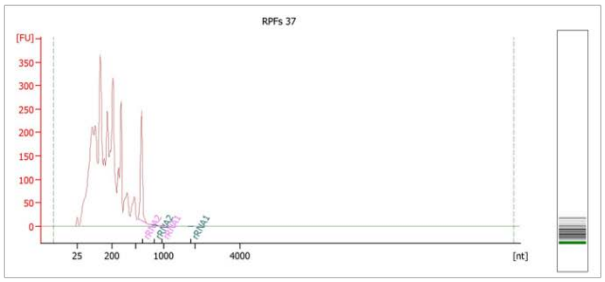
Paper V - Challenges and solutions for ribosome profiling with limited cell numbers: the example of murine male germ cells

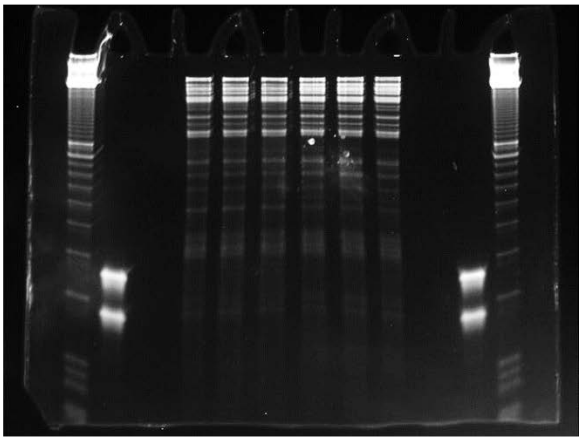
(in preparation)

Supplementary Figure 1. Effect of different digestion temperatures in ribosome footprinting.

RNA fragments are sensitive to variations in digestion temperatures, as seen by the decrease of RNA recovered when performing treatments at 37°C. These differences are not detectable by PAGE (right panel), however they can be measured using sensitive RNA-chips in an Agilent Bioanalyzer.







Temperature	RPFs (µg)
RT	7.75
37°C	2.59

Table S1. Estimation of amount of RNA expected to recover for each population from one Ho-FACS

					Expected ng/ FACS***					
		ng /10^6 cells**			100% lysate			75% RPFs + 25% total RNA		
	#Cells/FACS*	RPFs <200	RPFs >200	Total RNA	RPFs <200	RPFs >200	Total RNA	RPFs <200	RPFs >200	Total RNA
Spg	465944	1283.2	2177.4	9924.1	597.9	1014.6	4624.1	448.4	760.9	1156.0
Spc I	1666111	973.6	558.4	1081.5	1622.2	930.3	1801.8	1216.6	697.8	450.5
Spc II	825000	7840.0	850.9	1430.9	6468.0	702.0	1180.5	4851.0	526.5	295.1
Spd	9650000	56.6	59.8	135.7	546.6	576.9	1309.1	409.9	432.6	327.3
rSpd	6148000	63.9	47.2	373.0	393.0	290.3	2293.2	294.7	217.7	573.3
eSpd	1420000	287.9	658.6	1476.3	408.8	935.3	2096.3	306.6	701.5	524.1

* Average number of cells obtained for each population per FACS session

** Average amount of RNA recovered from 10^6 cells of each population

*** Estimated based on * and ** when using 100% or dividing cell lysates for transcriptome (25%) and ribosome profiling (75%)

Table S2. Summary statistics of ribo-seq data

	Sequencing depth	% reads mapped	% rRNA reads
RP_Ingolia	69M	66%	71.71%
RP_Kit_10ug	1.4M	21.46%	2.55%
RP_Kit 1ug	2.2M	47.02%	5.72%
RP_Kit_100ng	5.3M	28%	3.59%
RP_Kit_10ng	3.6M	8%	1.22%